U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Analysis of vegetation dynamics using time-series vegetation index data from Earth Observation Satellites

Arlete Silva Rodrigues

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto
Departamento de Geociências, Ambiente e Ordenamento do Território
Engenharia Geográfica
2014

Analysis of vegetation dynamics using time-series vegetation index data from Earth Observation Satellites

Arlete Silva Rodrigues
Doutoramento em Engenharia Geográfica
Departamento de Geociências, Ambiente e Ordenamento do Território
2014

Orientador
Dr. André Ribeiro da Silva de Almeida Marçal, Prof. Auxiliar
Departamento de Matemática, FCUP

Co-orientador
Dr. Mário Manuel de Miranda Furtado Campos Cunha, Prof. Auxiliar
Departamento de Geociências, Ambiente e Ordenamento do Território, FCUP

Acknowledgments

Completing the PhD and writing this thesis has been a long journey, but also an amazing and enriching experience in both professional and personal terms. I have deep gratitude for the many people who have helped me to build this dissertation over the last four years, and making this doctoral work an unforgettable experience.

First and foremost, I would like to give a special thanks to my supervisors, Dr. André Marçal and Dr. Mário Cunha, for their orientation and support. Our supervisor-student relationship began when I entered the M.S. program in Remote Sensing in 2007. Their support, schedule's flexibility, gentle encouragement and good advices contributed for a good working relationship and helped me to finish this thesis. My sincere thanks to both of you.

Another important supporter was my workroom colleague Dra. Isabel Pôças. Thank you for all your help, support, good conversations and pleasant lunch breaks.

I must also thank my friends (too many to list here but they know who they are!) for providing the support and friendship that I needed. A special thank you to my "weekend" reunion friends for making me laugh even in the not so-good moments.

Of course no acknowledgments would be complete without giving thanks to my family. My parents who work very hard during their entire life to provide the best for me and my brothers, and from who I never missed love and care. To my brothers who, although the distance, always gave me the best of the advices and support. Thanks for your care and our invigorating family reunions.

Last, but certainly not the least, I must acknowledge with tremendous and deep thanks to my life's partner, André Miranda. Our relationship began 12 years ago and since then I have been happier and more complete. I could not have complete this long journey without him by my side. His support, patient and unwavering belief in me, gave me the strength to finish my PhD dissertation and being here today. André was patient and extremely comprehensive in my worst moments, being capable of given me confidence and motivation. There are no words that can express my gratitude and appreciation for all you do and are for me. I love you.

I would like to thank the Fundação para a Ciência e a Tecnologia (FCT) for the research grant SFRH/BD/62189/2009, and also to the Dept. Geociências, Ambiente e Ordenamento do Território (DGAOT) and the Centro de Investigação em Ciências Geo-Espaciais (CICGE) for providing the resources and conditions needed for the development of my PhD work.

Abstract

This thesis focuses on the study of vegetation dynamics using satellite remote sensing data. Vegetation covers a substantial portion of the Earth's land surface, having an important role in the Earth's energy balance. Several factors, such as climate, abiotic environments or biotic interactions, and the dynamics of population and their activities have a great impact on the vegetation processes.

The study of the vegetation dynamics has been an active research topic since the late 19th century. Mapping vegetation is important and useful for natural resources management, land planning and/or environmental policymaking decision support. With the increasing availability of remotely sensed images, and the evolution of sensor technology (spectral, spatial, radiometric and temporal resolutions), vegetation mapping has become possible not only for local but also for global scales. The temporal resolution has a crucial impact in the field of study, as it provides data on different stages of the vegetation development. This advantage along with the use of long time-series periods lead to a more detailed and rigorous monitoring of the vegetation dynamics changes.

The main objective of this work consisted in the investigation and development of new methodologies to process and obtain information from satellite time-series data of vegetation. The research addressed two topics: crop monitoring and land cover classification. The investigations conducted on the topic of crop monitoring mainly resulted in the development of a software tool, PhenoSat, which is able to extract phenological metrics from satellite vegetation time-series data. PhenoSat obtains phenological information for the main growth stages and also solves two limitations found in other available software packages: (1) the impossibility to select an in-season region of interest (especially important when studying discontinuous canopies or dormancy period); and (2) the no-detection and/or no-extraction of double growth season information. PhenoSat is freely available at http://www.fc.up.pt/PhenoSat.

The investigations conducted on the land cover classification topic focused in two distinct regions: the African Continent and the Brazilian Amazon. Large areas and the presence of clouds are the main problems for land cover studies, due to the limitations on obtaining ground truth information for validation, and on using satellite data for classification in cloudy covered areas. To address these limitations, a new classification procedure is proposed based exclusively in satellite vegetation time-series.

Keywords: vegetation, land cover, classification, phenology, time-series, remote sensing, PhenoSat

Resumo

Esta tese foca-se no estudo da dinâmica da vegetação utilizando dados de detecção remota obtidos por satélite. A vegetação cobre uma parte considerável da superfície terrestre, desempenhando um papel importante no equilíbrio energético da Terra. Vários factores, como clima, ambiente abiótico ou interacções bióticas, e a dinâmica da população e suas actividades têm um grande impacto nos processos de vegetação.

O estudo da dinâmica da vegetação tem sido um tema de investigação frequente desde o final do século XIX. O mapeamento da vegetação é importante e útil para a gestão de recursos naturais, ordenamento do território e/ou apoio à decisão de políticas ambientais. Com a crescente disponibilidade de imagens adquiridas remotamente, e a evolução da tecnologia dos sensores (resoluções espectral, espacial, radiométrica e temporal), o mapeamento da vegetação tornou-se possível não apenas a nível local mas também global. A resolução temporal tem um impacto crucial na área de estudo, uma vez que fornece dados em diferentes fases do desenvolvimento da vegetação. Esta vantagem, combinada com a utilização de longos períodos de séries temporais, permite uma monitorização mais detalhada e rigorosa das alterações na dinâmica da vegetação.

O objectivo principal deste trabalho consistiu na investigação e desenvolvimento de novas metodologias para processar e obter informação a partir de dados de séries temporais de vegetação obtidas por satélite. A investigação abordou dois temas: monitorização de culturas agrícolas e classificação da cobertura do solo. As investigações realizadas sobre o tema da monitorização de culturas agrícolas resultaram principalmente no desenvolvimento de uma ferramenta computacional, PhenoSat, que é capaz de extrair métricas fenológicas a partir de séries temporais de vegetação obtidas por satélite. PhenoSat obtém informação fenológica para as principais fases de crescimento de culturas e resolve também duas limitações encontradas em outras ferramentas computacionais disponíveis: (1) a impossibilidade de seleccionar uma região de interesse durante o período de crescimento (especialmente importante em estudos de culturas com cobertos descontínuos e com períodos de dormência); e (2) a não-detecção e/ou não-extracção de informação sobre a fase de recrescimento. PhenoSat está disponível gratuitamente em http://www.fc.up.pt/PhenoSat.

As investigações realizadas sobre o tema da classificação da cobertura do solo focaram-se em duas regiões distintas: o Continente Africano e a Amazónia Brasileira. Grandes áreas e a presença de nuvens são os principais problemas para estudos de ocupação do solo, devido às limitações em obter dados de campo para validação, e em usar dados de satélites para classificação em áreas cobertas por nuvens. Para resolver estas limitações, é proposto um novo método de classificação baseado exclusivamente em séries temporais de vegetação obtidas por satélites.

Palavras-chave: vegetação, cobertura do solo, classificação, fenologia, séries temporais, detecção remota, PhenoSat

Contents

List of Figures and Tables

List of Abbreviations

ASTER – Advanced Spaceborne Thermal Emission and Reflection Radiometer

AVHRR – Advanced Very High Resolution Radiometer

EOS – Earth Observation Satellites

EVI – Enhanced Vegetation Index

MERIS – Medium Resolution Imaging Spectrometer

MODIS – Moderate Resolution Imaging Spectroradiometer

MVC – Maximum Value Composites

NDVI – Normalized Difference Vegetation Index

NIR – Near Infrared

NOAA – National Oceanic and Atmospheric Administration

PROBA – Project OnBoard Autonomy

SPOT – Satellite Pour l' Observation de la Terre

VI – Vegetation Indices

INTRODUCTION

1. Thesis Focus

This thesis focuses on the study of vegetation dynamics using time-series vegetation index data from Earth Observation Satellites (EOS).

Vegetation covers a substantial part of the Earth's surface (figure 1). Although 75% of the planet cover is an ocean of blue, the remaining 25% of Earth's surface is a dynamic green (NOAA, 2014). Vegetation is a dynamic component, strongly dependent of environmental conditions, having an important impact in the ecosystem (surface temperature, humidity, water balance, CO_2 concentration) and being indispensable for normal weather and climate.

Accurate and timely data that describes vegetation conditions is crucial to assess vegetation development and to best understanding and monitoring the land use practices. The field survey methods are generally associated to high costs, subjectivity and low spatial and temporal coverage, which limits the effectiveness of the process (Langley et al., 2001). The field collected data can be supplemented using EOS data. The capability to obtain regular observations at various scales, and information for areas with limited *in situ* accessibility, makes possible the use of remote sensing imagery and the study of vegetation from local to global scales.

EOS vegetation data has many applications, such as agriculture, forestry, ecology or urban planning. In this thesis, two main topics were addressed: land cover changes detection and crop monitoring.

Fig.1. The two sides of Earth depicting green vegetation, using Visible-Infrared Imager/Radiometer Suite data from April 2012 to April 2013. The darkest green areas represent abundant green vegetation, and the pale colours are sparse vegetation cover either due to snow, drought, rock or urban areas *(in NOAA, 2014*).

1.1. Thesis Structure

This thesis is divided in 5 main chapters and 4 appendices. The main chapters include an introduction, three research articles already published in scientific journals, and a section of conclusion remarks. The different chapters are organized as follows:

- Chapter I consists of a general introduction that summarizes and presents a brief state-of-the-art for the most important concepts addressed in this PhD work.

- Chapter II refers to the paper published in the IEEE Transactions on Geoscience and Remote Sensing journal, which presents a description of the PhenoSat software tool developed during this PhD work. This paper is a follow up of a paper presented in the MultiTemp2011 conference (appendix 1), where an initial presentation of the PhenoSat software was made. PhenoSat permits to extract phenological information from satellite vegetation index time-series data using a simple and user-friendly interface. As PhenoSat is freely available for research, technical and educational purposes, good feedback has been received, permitting to improve the functionalities of this tool. The complementary work related to the development of PhenoSat resulted in two conference proceedings papers, included in appendices 1 and 2, as well as a Book Chapter (appendix 3).

- Chapter III summarizes the research work developed for the creation of a new methodology of classification based exclusively in time-series of vegetation data. The identification of potential land cover changes using the methodology proposed was tested in the African Continent. This work resulted in a paper published at the International Journal of Remote Sensing.

- Chapter IV expands the research done in the previous chapter. The classification procedure was applied in Rondonia, a region of the Brazilian Amazon, and the research work was initially presented at the 33rd Canadian Symposium on Remote Sensing (appendix 4). The large number of land cover types and the similarity between the vegetation index profiles affect the classification results. Thus, the effects of the similarity were tested using new approaches to reduce the number of classes

and to evaluate the variability over the years. This research work resulted in a paper published in the Canadian Journal of Remote Sensing.

- Chapter V presents the conclusions section, which summarizes the main achievements obtained with this PhD project, and also some suggestions for future work.

1.2. List of Publications

The following thesis includes seven research articles. Three of them have been published in scientific journals during the duration of this PhD work, three were presented in international conferences and one of them resulted from an invitation for a chapter in the Springer EARSeL Book 'Multitemporal Remote Sensing: Method and Applications'. The articles published in scientific journals are part of this thesis (chapters II, III and IV) and the others are included as appendices. A reference list of these papers is given below:

Scientific Journals:

- Arlete Rodrigues, André R.S. Marçal, and Mário Cunha (2013). Monitoring vegetation dynamics inferred by satellite data using the PhenoSat tool*. IEEE Transactions on Geoscience and Remote Sensing, 51,* 2096-2104

- Arlete Rodrigues, André R.S. Marçal, and Mário Cunha (2013). Identification of potential land cover changes on a continental scale using NDVI time-series from SPOT VEGETATION. *International Journal of Remote Sensing, 34,* 8028-8050

- A. Rodrigues, A.R.S. Marçal, D. Furlan, M.V. Ballester, and M. Cunha (2013). Land cover map production for Brazilian Amazon using NDVI SPOT VEGETATION time series. *Canadian Journal of Remote Sensing, 39,* 277-289

Conference Proceedings:

- Arlete Rodrigues, André R.S. Marçal, and Mário Cunha (2011). PhenoSat - A tool for vegetation temporal analysis from satellite image data. *Proceedings of the 6th International Workshop on the Analysis of Muti-temporal Remote Sensing Images (Multi-Temp), at Trento, Italy, 12-14 July*

- Deise Furlan, Arlete Rodrigues, Mário Cunha, André R.S. Marçal, and Maria Victoria Ballester (2012). Land cover classification in Rondonia (Amazonia) using NDVI time

series data from SPOT-VEGETATION. *Proceedings of the 33rd Canadian Symposium on Remote Sensing, at Ottawa, Canada, 11-14 June*

- Arlete Rodrigues, André R.S. Marçal, and Mário Cunha (2012). Phenology parameter extraction from time-series of satellite vegetation index data using PhenoSat. *Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, at Munich, Germany, 22-27 July*

Book Chapter:

- Arlete Rodrigues, André R.S. Marçal, and Mário Cunha (*submitted*). PhenoSat - A tool for remote sensing based analysis of vegetation dynamics. *In Springer EARSeL Book 'Multitemporal Remote Sensing: Method and Applications'*

2. Analysis of Vegetation Dynamics using EOS Data

2.1. Earth Observation Satellites

The first images from space were taken on the sub-orbital V-2 rocket flight launched by the U.S. on October 24, 1946 (Reichhardt, 2006). From 1946 to 1952, upper-atmosphere research was conducted using V-2s and Aerobee rockets. In 1954, an Aerobee sounding rocket photographed an unknown tropical storm in the Gulf of Mexico (Hubert and Berg, 1955). Despite these early spaceflights, the space age is generally considered to have begun with the historic launch of the first artificial satellite, Sputnik 1 (figure 2), on October 4, 1957, by the Soviet Union (NASA History Program Office, 2014). A satellite is a moon, planet or machine that orbits a planet or star. Usually, the word “satellite” refers to a machine that is launched into the space and moves around Earth or another body in space.

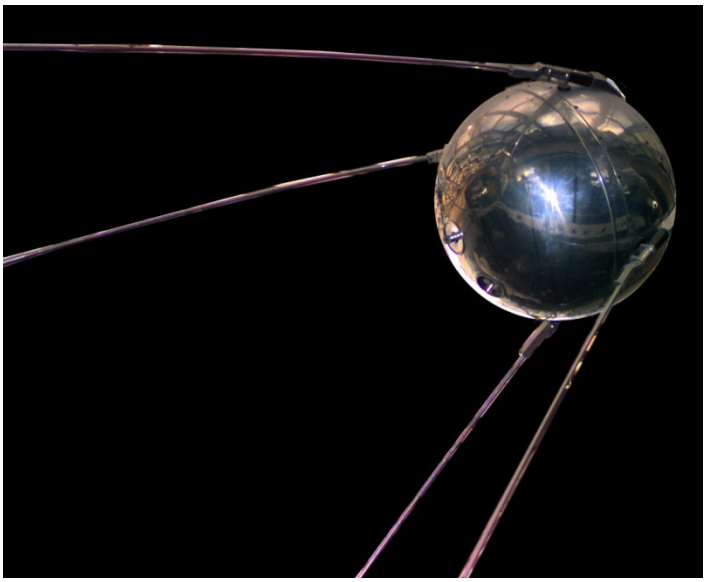


Fig. 2. Sputnik 1, launched from Baikonur, USSR, on October 4, 1957.

Since the launch of the first artificial satellite, Sputnik 1, thousands of satellites have been launched into orbit around the Earth. The largest artificial satellite currently orbiting the Earth is the International Space Station (ISS). The ISS (figure 3) is a collaboration of 15 nations working together to create a world-class, state-of-the-art orbiting research facility (NASA Missions, 2014).

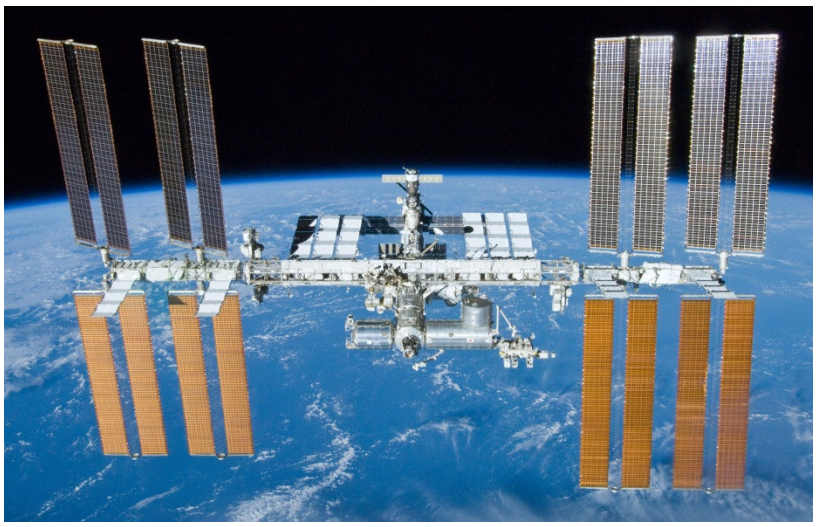


Fig. 3. International Space Station, orbiting Earth since 1998.

Earth Observation Satellites (EOS) are artificial satellites specifically designed to observe the Earth from orbit, and acquire global observations of the land surface, biosphere, solid Earth, atmosphere and oceans. Satellites orbit Earth at different heights, speeds and along different paths. The Earth observation techniques are used to collect information about physical, chemical and biological systems of the planet Earth. This information permits to assess and monitor the status and changes in the Earth's environment, being important to improve the social and economic well-being. The process of gathering information from Earth's surface has become more and more sophisticated due the technological advances in the surveying techniques and the improvements in the analyses and processing methodologies.

The U.S. Explorer VI made the first satellite (orbital) photographs of Earth on August 14, 1959. In April 1960, the polar-orbiting Television Infrared Observing Satellite (TIROS)-1 was designed to provide a view from space of Earth's cloud patterns (NASA Science Missions, 2014). Later versions of these satellites added microwave instruments to provide estimates of temperature, pressure, and humidity, which are essential to weather forecast models. Several improvements were made in the TIROS series (now called Polar-orbiting Operational Environmental Satellites (POES)), making these satellite measurements indispensable around the world.

In the early 1970's the NASA created the Landsat program, the largest program for acquisition of imagery of Earth from space (Mack, 1990). Landsat represents the world's longest continuously (four decades) acquired collection of space-based moderate-resolution land remote sensing data (figure 4). The first three satellites were using 80-

meter resolution multispectral scanners (MSSs), providing images of large areas of the Earth's surface in four different color bands since 1972. The Landsat mission was extended to Landsat 4 (1982) and 5 (1984), which collected information in seven spectral bands through an enhanced 30-meter sensor called the thematic mapper (TM). The Landsat 6 failed; Landsat 7 (Enhanced Thematic Mapper Plus, ETM^{+}) and Landsat 8 (formerly the Landsat Data Continuity Mission, LDCM) were launched in 1999 and 2013, respectively, and continue in orbit and supplying data.

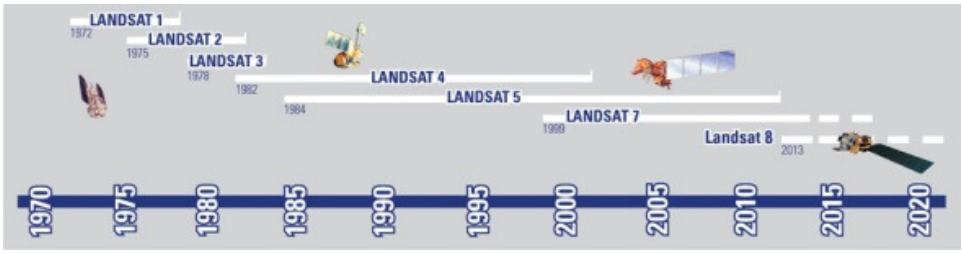


Fig. 4. Landsat Missions Timeline (*in* http://landsat.usgs.gov/about_mission_history.php).

During the last decades, several satellites were created and launched to collect useful Earth's surface data. EOS images are an important source of information in a wide range of applications in different fields, such as ocean, meteorology, geology, forestry, landscape and regional planning. Depending on the function and orbit, each satellite instruments have different spatial, spectral, radiometric and temporal resolutions.

The spatial resolution is related with the minimum size of detail observable in an image, defining the pixel size of the images covering the Earth surface. Very high-resolution (0.82 – 4 meters) EOS images became available in the autumn of 1999, with the launch of the first IKONOS satellite (SIC, 2014). Since then, different very high-resolution satellites were designed and launched, providing information of the Earth's surface with an amazing detail.

The spectral resolution describes the ability of a sensor to distinguish features in the electromagnetic spectrum (Campbell, 2002). The radiometric resolution refers to the sensitivity of the sensor to variations in brightness and denotes the number of grayscale levels that can be imaged by a sensor. The higher the radiometric resolution the more sensitive it is to detecting small differences in reflected or emitted energy. Finally, the temporal resolution specifies the revisiting frequency of a satellite sensor for a specific location.

The concept of temporal resolution is one of the most important elements for several remote sensing applications. It is crucial to monitor Earth's surface changes or short time-varying phenomena, naturally occurred or caused by humans. The time factor is an important aspect particularly in the tropics, where persistent clouds limit the views of the Earth´s surface.

In general, the revisiting period of a satellite sensor is several days (e.g. 16-days for Landsat 8). However, some satellite sensors are able to image the same area of the Earth in periods from one to five days (e.g. IKONOS, Satellite Pour l'Observation de la Terre (SPOT) 5, Moderate Resolution Imaging Spectrometer (MODIS), Project for On-Board Autonomy V (PROBA V), Medium Resolution Imaging Spectrometer (MERIS)). The temporal resolution of a satellite sensor depends on a variety of factors, including satellite/sensor capabilities, swath overlap and latitude.

The number of satellite sensors capable to acquire high resolution temporal data has increased during the last years. Most of these sensors are capable to acquire time-series of vegetation data. Vegetation has unique spectral signatures, which evolve with the plant vegetative cycle. Thus, vegetation mapping through quantifying vegetation cover presents valuable information for understanding the natural and man-made environments, from local to global scales. Furthermore, the potential of remote sensing vegetation data in agriculture and crop management is high as the multispectral reflectance and temperatures of crop canopies are related to photosynthesis and evapotranspiration (Basso et al., 2004).

2.2. Vegetation Indices

Vegetation indices (VI) combine information from different wavelengths, particularly in the RED and near-infrared (NIR) portions of the spectrum, to enhance the vegetation signal. Such indices allow reliable spatial and temporal inter-comparisons of terrestrial photosynthetic activity and canopy structural variations (Huete et al., 2002). They are generally computed for all pixels in time and space, regardless of biome type, land cover condition and soil type, thus representing true surface measurements. Due to their simplicity and ease of use, VI have a wide range of applications, being an important tool to assess how environmental changes affect the distribution and dynamics of vegetation, particularly at large temporal and spatial scales, and/or in areas of limited *in situ* data.

Unfortunately, the VI data collected using EOS sensors can be affected by diverse factors such as illumination and viewing geometry, clouds, aerosol, shadow and water vapour (Tucker et al., 1985; Cihlar et al., 1994), which could generate some noisy values. The poor quality of observations can affect the subsequent data analysis and can limit the use of these datasets (Mithal et al., 2011). The Maximum Value Composites (MVC) process can help deal with this problem, removing a significant part of that noise. It consists on analyzing the VI values on a pixel-by-pixel basis, in a predefined time-period, and retaining the highest value for each pixel location (Holben, 2007). The MVC imagery is obtained when all pixels have been evaluated and is highly related to the green vegetation dynamics.

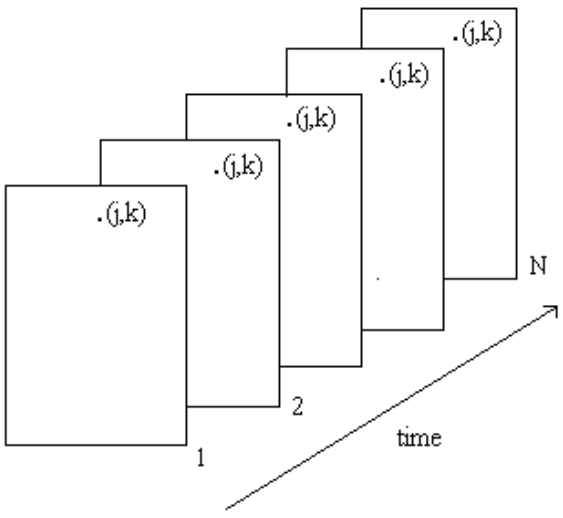


Fig. 5. EOS vegetation indices data organized in a three dimensional data cube.

EOS VI data are easily available and can be searched and ordered with reduced or no cost over the internet. These datasets are organized in two-dimensional images, with the time acting as a third dimension in the data cube (figure 5). For an observation *t*, of pixel location *(j,k)*, $N_t(j,k)$ gives the VI value at time *t* (Rouse et al., 1974). The set of observations $N_{ti}(j,k)$, with *i=1,..,N*, is the VI time-series for pixel *(j,k).*

Several VI have been developed for different sensors and with different purposes, such as Perpendicular Vegetation Index- PVI (Richardson and Wiegand, 1977), Green Difference Vegetation Index- GDVI (Sripada et al., 2006), Soil Adjusted Vegetation Index- SAVI (Huete, 1988), Vegetation Condition Index- VCI (Kogan, 1995), Leaf Area Index- LAI (Watson, 1947), Normalized Difference Vegetation Index- NDVI (Rouse et al., 1974) and Enhanced Vegetation Index- EVI (Huete et al., 1997).

NDVI is the most widely used vegetation index to quantify the vegetation condition, being directly related to the photosynthetic capacity and energy absorption of plant canopies (Sellers, 1985). It is determined using the RED and NIR bands of a given image and is expressed as Eq.1 (Rouse et al., 1974).

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (1)$$

The NDVI values range from -1 to +1. Due to light absorption by chlorophyll, the RED spectrum reflection is always lower than in the NIR for the green vegetation. Thus, NDVI values for the vegetation cannot be lower than 0. In a practical sense, the values that are equal or below 0.1 correspond to bare ground and water bodies due their very low reflectance in the NIR band. Values comprising the [0.2 – 0.5] range are related with sparse vegetation of grassland and shrubs, and NDVI values near to 1 indicate the highest possible density of green leaves (dense vegetation as the closed evergreen tropical forest) (Weier and Herring, 2000).

The choice of NDVI to monitor the changes in vegetation phenology (growing, flowering, harvesting, and senescence) during the seasons and from year to year is advantageous, as it is easily computed without assumptions regarding land cover classes, soil type or climatic conditions. However, the NDVI index presents some disadvantages, such as the: inherent nonlinearity, additive noise effects, saturated signals over high biomass conditions and is very sensible to canopy background brightness (Vermote and Vermeulen, 1999; Vermote et al., 2002).

The main function of VI, other than NDVI, is to compensate for the effects of disturbing factors on the relationships between vegetation spectral reflectance as measured by crop characteristics (crop type, canopy biomass, leaf area index) (Bouman, 1995). To mitigate the effects of soil brightness in cases where vegetation is sparse, some distance-based VI (e.g. PVI and SAVI) have been developed, being particularly important in arid and semi-arid environments, where pixels contain a mixture of green vegetation and soil background (Huete and Jackson, 1988).

The EVI index was developed by the MODIS science team to take full advantages of the sensor capabilities. It was designed to enhance the vegetation signal with improved sensitivity to high biomass regions and improved vegetation monitoring through a de-coupling canopy background signal and a reduction in atmosphere influences (Huete et al., 1997). EVI is defined as Eq.2:

$$EVI = G \times \frac{\rho_{NIR} - \rho_{RED}}{(\rho_{NIR} + C_{1} \times \rho_{RED} - C_{2} \times \rho_{BLUE} + L)} \quad (2)$$

where $\rho_{NIR}/\rho_{RED}/\rho_{BLUE}$ are atmospherically-corrected or partially atmosphere corrected (Rayleigh and ozone absorption) surface reflectances, G is the gain factor, L is the canopy background adjustment that addresses non-linear, differential NIR and RED radiant transfer through a canopy, and $C1$, $C2$ are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band. The correction of aerosol impact on the final index makes use of reflectance measurements within visible blue, which is not always available (case of Advanced Very High Resolution Radiometer (AVHRR)).

In comparison with NDVI, which is only chlorophyll sensitive, EVI is more linearly correlated with green leaf area index, canopy type, plant physiology and canopy architecture (Boegh et al., 2002), less prone to saturation in temperate and tropical forests (Huete et al., 2006; Xiao et al., 2004a) and minimally sensitive to residual aerosol contamination (Miura et al., 1998; Xiao at al., 2003). Many studies found that NDVI becomes saturated over highly vegetated areas and does not respond to variation of NIR reflectance when the red reflectance is low (Gitelson, 2004; Wardlow et al., 2007). To compensate for the effects of NDVI saturation over high biomass areas, EVI tends to present relatively low values in all biomes and also lower ranges over semi-arid sites (Jiang et al., 2008).

As an illustration, figure 6 presents the EVI and NDVI 16-days composites, acquired by TERRA MODIS sensor in a vineyard (Douro region, Portugal), for the 2003-2008 years. EVI presents lower values than NDVI throughout. The soil vegetation growth during the winter season causes higher variations in the NDVI profile, which is less observable in the EVI data, presenting a smoother profile and consequently a more clear identification of the vineyard growing season.

Among the several VI, the NDVI and EVI indices provide detailed insight into biotic activities (Running et al., 1994) and have been applied in a broad range of studies such as global climate change, phenological and crop growth monitoring, yield prediction, climatic and biogeochemical modelling, to name a few (e.g. Justice et al., 1998, 2002; Running et al., 2004; Cunha et al., 2010). The two vegetation indices, EVI and NDVI, complement each other in global vegetation studies and improve upon the detection of vegetation changes and extraction of canopy biophysical parameters (Huete et al., 1997).

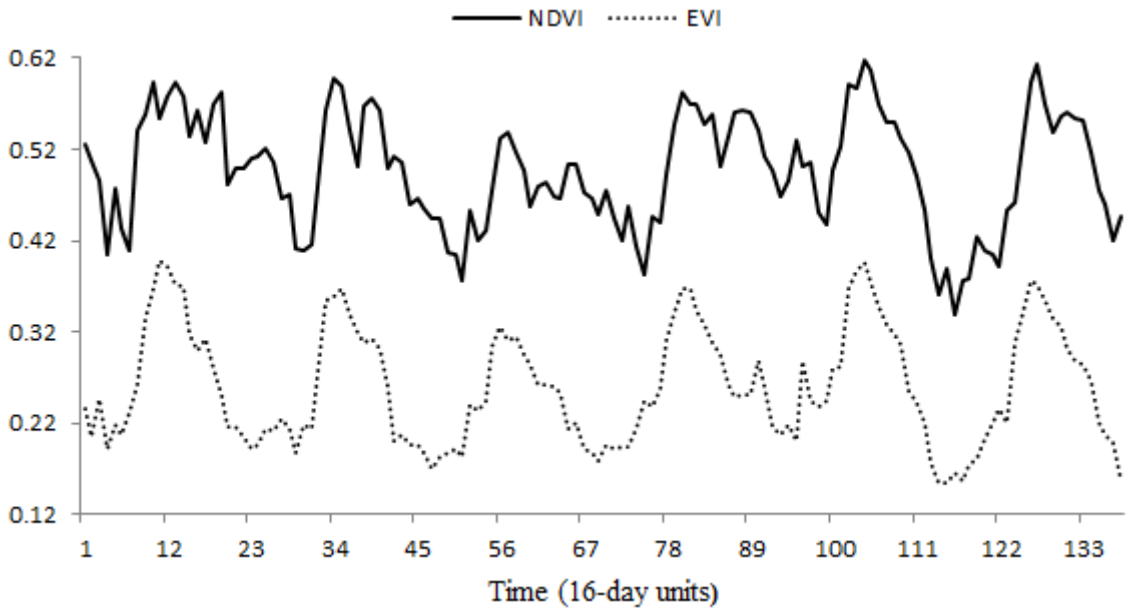


Fig. 6. Example of temporal series of EVI (dotted line) and NDVI (solid line) acquired from TERRA MODIS sensor, for a vineyard region in Douro (Portugal) between 2003 and 2008.

2.3. Vegetation Dynamics using Remotely Sensed Data

Vegetation phenology is the science of studying life-cycle events of plants, and their responses to seasonal and inter-annual variation in climate (Morisette et al., 2009). The development of phenological models has permitted to study the relationship between biological events and the environment. The main methods used to monitor the phenology of a region are by field observations, bioclimatic models (e.g. Chuine, 2000), airborne pollen data (e.g. Helbig et al., 2004) or, more recently, through the analysis of remote sensing data. Remote sensing phenological studies can be done at regional or global scale, which is not feasible using field phenological studies, due the temporal and spatial limitations (Moulin et al., 1997; Reed et al., 1994; Zhang et al., 2003).

Recent technological advances have resulted in a new field of phenological research concerned with the observation of whole ecosystems and stands of vegetation on global scales. The most successful approach is based on satellite VI time-series data to track the growth stages of green vegetation (emergence, growth, maturity, and harvest). Information about the time of occurrence of these stages was considered as indicators revealing the characteristics of the vegetation in the surface (Schwartz and Karl, 1990). The VI are useful indicators of vegetation properties and greenness in the spatial and temporal domain. Thus, the analysis of the VI time-series through time permits the extraction of appropriate metrics to describe vegetation dynamics, allowing a better monitoring and understanding of the biophysical changes in the vegetation cover and phenology in different ecosystems (Bradley and Mustard 2008).

The study of vegetation dynamics at continental to global scales was enabled in 1981 by the AVHRR onboard National Oceanic and Atmospheric Administration (NOAA) satellite (Tucker et al., 1985; Eidenshink, 1992). AVHRR has a long enough temporal global coverage of EOS NDVI products for vegetation condition and ecosystem studies in global scales. These data series are useful for learning from past events, monitoring of current conditions and studying the trend of NDVI for detection of long-term vegetation dynamics (van Leeuwen et al., 2006).

The main advantages of AVHRR are the long period of collections, high radiometric resolution and the temporal ability for tracking fast the surface coverage changes. The spatial resolution of 1.1km lends itself to observing surface conditions on a regional and continental scale. However, this spatial resolution can be a disadvantage if the area of

interest is relatively small. Furthermore, AVHRR sensors were not originally designed for vegetation monitoring (Teillet at al., 1997). The sensors are sensitive to water vapor in the atmosphere due to the wide spectral bands, particularly the NIR channel, which clearly affects the calculated NDVI value. The AVHRR lacks in terms of quality and atmospheric corrections can be partially solved using modern sensors, such as SPOT VEGETATION, MODIS, MERIS, and PROBA V. These newer instruments are equipped with better navigation, atmospheric correction and improved radiometric sensitivity systems (Gobron et al., 2000).

Long time-series of VI data can provide information on shifts in the spatial distribution of bio-climatic zones, indicating variations in large-scale circulation patterns or land-use changes. There is thus a great interest in maintaining data continuity and compatibility across the sensor datasets (Gitelson and Kaufman, 1998). Compatibility problems among the satellite data products exist due to the differences in sensor characteristics and algorithms used to process the data. The purpose of the VI continuity is to normalize the spectral, spatial, temporal and radiometric differences in order to extend stable temporal datasets to monitor the ecosystem changes. The VI continuity/compatibility is obtained when VI values computed from different sensors become the same for the same target under identical conditions (Yoshioka et al., 2002).

Despite the great importance of the remote sensing data in phenological studies (Bin et al., 2008; Pettorelli et al., 2005) and the widespread access to satellite vegetation data, there is still a gap between data and the required information. The large amounts of data and the inconsistency of image quality can make the analysis and extraction of relevant vegetation information a difficult and time-consuming process (Badeck et al., 2004).

The VI time-series obtained by EOS images contain a combination of phenological (inter-season) and trend changes (intra-seasons), in addition to various noise components such as atmospheric disturbances, viewing and solar illumination variability, cloud cover and others. In order to improve the image quality, several approaches have been applied to filter noise and optimize the satellite signal. The creation of composites permits to reduce the atmospheric disturbances and obtain a higher percentage of clear-sky data. However, the MVC process is not generally sufficient to eliminate all unrealistic variability from VI time-series, such as residual effects of sub-pixel clouds and prolonged cloudiness elements (Jonsson and Eklundh, 2004). If the composite period is long, the land surface does not remain static; and if the composite is too short, the atmospheric disturbance cannot be removed efficiently, particularly in cloudy regions. Furthermore, additional noise

may be also introduced by the process of overlaying several images (for example due to image registration). These negative effects require further processing in the form of data smoothing.

The development of robust, accurate and fast filtering models, being capable of smoothing the data without introducing artifacts or suppressing natural variations of the vegetation, has been an important research topic. The difficulty/lack in obtaining reference (ground-based) data for validation, as the coarse scale of the sensors and the required representativeness of the reference measurements in time and space, represent the main difficulties of this subject (Justice et al., 2000).

Table 1. Filtering methods proposed for smoothing remotely sensed time-series of vegetation indices.

Filtering Method	***Some Prominent Applications***
Running Medians	Velleman (1980)
Best Index Slope Extraction	Viovy et al. (1992); Lovell and Graetz (2001)
Weighted Least Squares Windowed Regression	Swets et al. (1999)
Harmonic Series and Higher Order Splines	Roerink et al. (2000); McCloy and Lucht (2004); Bradley et al. (2007)
Wavelets	Li and Kafatos (2000); Sakamoto et al. (2005)
Asymmetric Gaussian	Jonsson and Eklundh (2002);
Double logistic	Zhang et al. (2003); Beck et al. (2006)
Savitzky-Golay	Chen et al. (2004)
Mean Value Iteration	Ma and Veroustraete (2006)
Whittaker smoother	Atzberger and Rembold (2009); Atzberger and Eilers (2010)

During the last years, different filtering techniques have been proposed (summarized in Table 1). In general, data smoothing facilitates the satellite time-series analyses, by eliminating the unrealistic abrupt peaks and aberrant values that appears in the VI profile (Fontana et al., 2008). Moreover, it permits a better observation of the vegetation changes over time and the identification of the main and double growing seasons, which is not always clearly possible using the VI original data. This is illustrated in the example presented in figure 7.

Many studies have focused on VI noise reduction techniques comparison, but there is not yet an agreement on which filter is the best to smooth the remotely sensed time-series (Chen et al., 2004; Beck et al., 2006; Hird and McDermid, 2009).

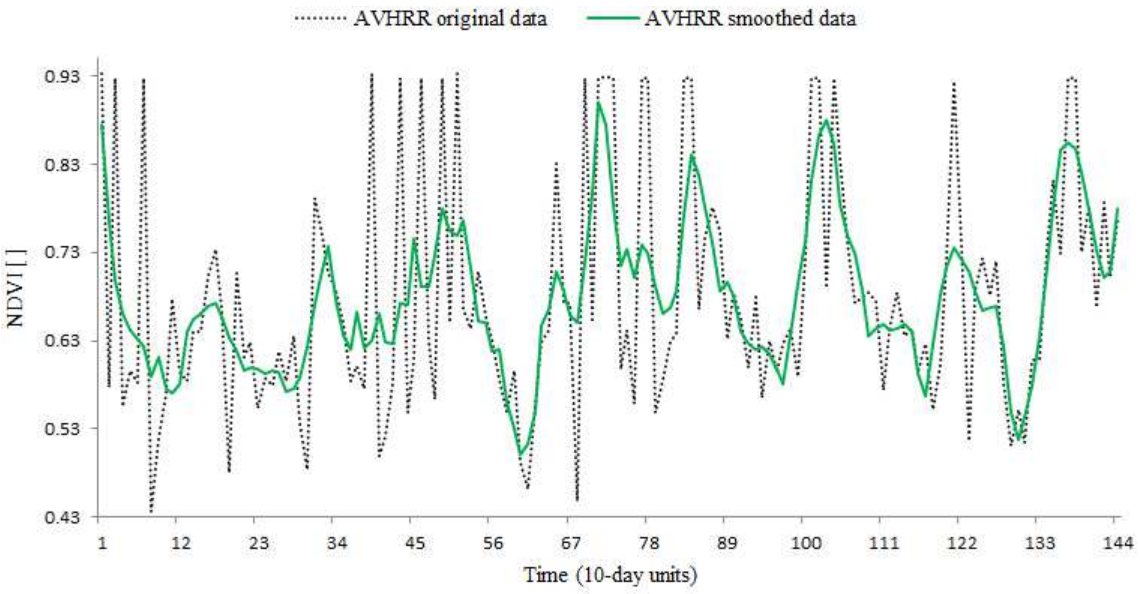


Fig. 7. Example of a temporal series of NDVI acquired from NOAA AVHRR sensor, for a semi-natural meadows region in Montalegre (Portugal) between the years 2001 and 2004. The dotted black line corresponds to the original NDVI data and the solid green line the smoothed data produced using the Savitzky-Golay method.

2.4. Crop Monitoring

Agricultural production strategies have changed dramatically over the past two decades. Many of these changes have been driven by human population increase, urbanisation, economic and political decisions (Reddy et al., 2003). The use of large machineries and applications of chemicals and fertilizers may contribute to negative environmental applications, such as soil fertility deterioration or water pollution. There is thus the need of an efficient approach to monitor the natural resources ensuring the quality, nutritional value, sustainability, and safety of agricultural products. Technological investments for crop management have become one of the most important approaches in the last twenty years, and particularly in the 21st century.

Crop management aims increasing the efficiency of resources use and reducing the uncertainty of decisions required to manage variability on farms (NRC, 1997). It is an integrated, information and production-based farming system that is designed to increase the efficiency, productivity and profitability of farm production. The crop management data generally include: crop and soil information, microclimatic data, surface and subsurface drainage conditions, fertilization planning, control of plagues/diseases and water availability (Venkataratnam, 2001). Due the impacts of the crop management in the environmental, social and economic aspects, various research and technological investments have been developed during the last two decades (Schellberg et al., 2008). In order to collect and utilize the information efficiently, a vast array of tools, including software, hardware and management practices are available, such as Global Positioning System (GPS) receivers, yield monitoring and mapping, Geographic Information Systems (GIS) and remote sensing.

The applications of remote sensing technologies in crop monitoring are increasing rapidly due to the great improvements made in the spatial, radiometric, spectral and temporal satellite resolutions. Besides providing a spatial overview of the land surface, the satellite remote sensing data provide spectral reflection information, which can be measured and used to monitor phenology, stage type and crops health (pionnering studies were made by Allen et al., 1969; Gausman et al., 1969; Wooley, 1971; Allen et al., 1973; Gausman and Allen, 1973; Gausman and Hart, 1974 and Gausman et al., 1974). Furthermore, the ability of some satellite sensors to collect imagery at frequent time intervals is an important element for crop monitoring. Multitemporal images permit to observe how the vegetation

changes throughout the growing season, and better monitor the changes naturally occurring or induced by humans.

Vegetation phenology based on remote sensing data refers to the spatio-temporal development of the vegetated land surface as revealed by satellite sensors (de Beurs and Henebry, 2004). The main assumption behind all methods for phenological determination from satellite sensor data is that the signal is related to measures of vegetation. A time-series of a given VI follows annual cycles of growth and decline. Thus, deriving phenological metrics from remotely sensed data consists on the analysis of the seasonal VI trajectory, and identifying critical points such as the start of the season or the end of season. However, this is not a straightforward process. Phenological metrics exploit the seasonal growth cycle information, which could be influenced by non-climatic factors, biogenic and anthropogenic disturbances (fires, land degradation, insect attacks), or temperature and rainfall variations (Julien and Sobrino, 2009; Potter et al., 2003).

Phenological metrics can be derived from satellite VI data in several ways. Some researchers use complex mathematical models, others employ thresholds. The most common method is based on thresholds, assuming that the growing season has started when filtered or partially smoothed VI values exceed a given value (Lloyd, 1990; Reed et al., 1994; Alberte, 1994; Zhou et al., 2003). Instead of using thresholds, other methods, as derivatives or fitting models, can also be applied to retrieve phenological information from EOS VI data. Each method has its own advantages and disadvantages (Hird and McDermid, 2009; de Beurs and Henebry, 2010; Viovy et al., 1992). However, the fitting models can be easily adapted to a wide range of situations, being the most chosen category in the last years.

The first method of the vegetation phenology based on EOS data using the fitting model was presented by Badhwar in 1984, and since then several methods have been developed (e.g. Jonsson and Eklundh, 2002; Zhang et al., 2004). Beck et al. (2006) and Hird and McDermid (2009) compared some of the fitting algorithms available in the literature and conclude that double logistic and asymmetric Gaussian functions are the most suitable for describing the vegetation dynamics, with a slightly better performance and flexibility for the double logistic.

Developing algorithms to automatically remove the time-series noise and retrieve land surface phenology metrics from satellite data has been a popular research topic for the last decade. TIMESAT (Jonsson and Eklundh, 2004) is the most known software for time-series phenology analysis, being used in several research studies (e.g. Gao et al., 2008;

Verbesselt et al., 2012; Zeng et al., 2013). It is an open source software and provides three different smoothing functions to fit the time-series data: asymmetric Gaussian, double logistic and adaptive Savitzky-Golay filter. TIMESAT uses a simple method, based on thresholds, to determine a set of phenological metrics, including the start of season, mid-season and end of season. This method works well in ecosystems with predictable minimum and maximum VI values, however it can not adapt so well to different vegetation dynamics over the years, caused by uncontrollable conditions (drought year, unseasonal snow, fire, plagues and diseases), and crops with partial ground cover or period of dormancy.

Besides TIMESAT, there are other software packages allowing the analysis of the satellite time-series, reduction of noise components and/or extraction of phenological metrics from satellite time-series data. HANTS (Roerink et al., 2000), TiSeG (Colditz et al., 2008), TSPT (Prados et al., 2006; McKellip et al., 2008), PPET (McKellip et al., 2010), TIMESTATS (Udelhoven, 2011) and Enhanced TIMESAT (Tan et al., 2011) are some examples. Although these software products have important functionalities for the extraction of phenological information, they present two great limitations: the first is related with the extraction of information for the double growth-season, and the second is the impossibility to define an in-season temporal region of interest that is important for crops with a dormancy period. Moreover, they were tested only in a short number of vegetation types and detailed comparisons with field data are still scarce.

Based on the increasing importance of phenological research for the management resources and yield production, this PhD work had as its main purpose the creation of a new software package (PhenoSat) to extract phenological information from satellite VI time-series. PhenoSat permits to obtain information for the main growth vegetation stages and solves the main limitations found in the aforementioned software packages. Figure 8 shows a schematic procedure for the calculation of PhenoSat derived phenological metrics. The NDVI original values were obtained from SPOT VEGETATION sensor for a vineyard crop in Portugal, for the year 2000.

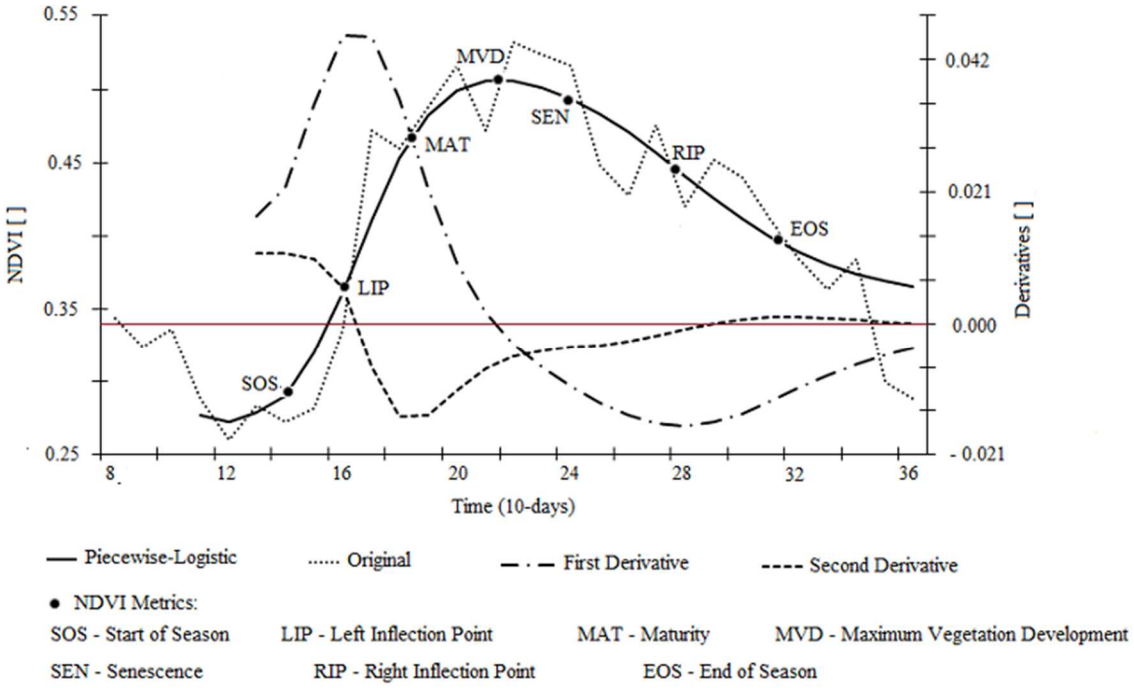


Fig.8. Representation of PhenoSat derived phenological stages, using the maxima and minima of the first and second derivatives of the NDVI fitted data.

2.5. Land Cover Change Detection

Land use and land cover is essential for many planning and management activities concerned with the Earth's surface, because it constitutes a key environmental information for a variety of scientific, resource management and policy purposes, as well as for a range of human activities (Cihlar, 2000).

Classifying and mapping vegetation is an important task for monitoring natural resources as vegetation provides a base for all living beings and plays an essential role in affecting global climate change (Xiao et al., 2004b). Traditional methods (e.g. field surveys, literature reviews, map interpretation and collateral and ancillary data analyses) are not effective to acquire vegetation covers because they are time-consuming, date lagged and often too expensive. The technology of remote sensing offers a practical and economical means to study vegetation cover changes, especially for large areas (Langley et al., 2001). The use of remote sensing in land use/land cover mapping is one of the most important applications of modern satellite sensor technology, as it provides data from which updated land cover map information can be extracted efficiently, at reduced costs.

The EOS data frequently used for land cover mapping are mostly high spatial resolution image data. The high spatial resolution multispectral data, provided by Landsat or SPOT High Resolution Visible (HRV) sensors, have been widely used to produce a large variety of land cover maps (Aban et al., 2002; Kokalj and Oštir, 2007; Marçal et al, 2005; Poças et al., 2011). Over the years, the improvements on sensors instruments leading to the production of more accurate land cover maps, due the very high spatial (e.g. IKONOS) and spectral (e.g. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)) resolutions. The automatic classification based on multispectral images usually uses high spatial resolution images (a single one or, a few from different seasons in a year) with multiple spectral bands. However, to detect with precision all the seasonal dynamics and the changes in the land cover, this type of classification is not appropriate. Multiple measurements over time are necessary to capture all the changes occurring naturally or caused by different factors (e.g. fire, human interventions or climate changes). The availability of satellite sensors with high temporal resolution (image per day or per 2-3 day periods) allows the acquisition of image sequences in time (time-series), which permits the accurate monitoring of the seasonal vegetation development as well as the detection and assessment of land cover changes (Lunetta et al., 2006). Besides that, these temporal data

can be an alternative to standard classification approaches, particularly in tropical regions where the presence of cloud cover seriously limits the use of passive satellites (Carreiras et al., 2003).

The automatic classification based on satellite time-series needs to be slightly different from the standard approach. The multispectral image classification usually uses high spatial resolution images with multiple spectral bands, and focuses on finding patterns on the spectral response having into account known land cover groups. Instead of using the multispectral values as features for each pixel, the temporal image classification uses the temporal yearly profiles as features. This type of classification is based on image sequences acquired over a period of time, thus providing high frequency observation of the land surface.

Since 1981, the AVHRR sensor has made the monitoring of the land surface processes (Hansen et al., 2000, Loveland et al., 2000). The long period of AVHRR data record is very valuable for global land cover change studies (Hansen and DeFries, 2004). However, the AVHRR data, as mentioned, have several limitations such as insufficient geometric correction, unsatisfactory atmospheric correction, or cloud masking (Cihlar et al., 1997; Roy, 2000). The MVC process can improve the AVHRR data by reducing the atmospheric effects, but some difficulties arise when all the entire compositing period is cloudy. The new era of temporal EOS data, well-calibrated and corrected geometrically, such as SPOT VEGETATION, MODIS and MERIS, provide the possibility of consistent long-term observations and trend analysis, important for multiple global-change studies.

The access to satellite temporal data is currently widespread, at reduced or no cost. The improvements on data processing made possible to extract data for land cover mapping at continental and global scales. However, there are some problems (diversity of inputs, labor to process the remote sensing data, difficulty in collect ground truth data) that make the production and update of land cover maps at these scales a difficult and a lengthy process (Loveland et al., 2000; Muchoney et al., 1996).

One of the main objectives of this PhD work was to investigate and develop a new methodology that allows the detection of potential land cover changes using a classification method based exclusively in VI time-series data. Similarity measures are used to perform a hierarchical aggregation of the land cover vegetation types, level-by-level, until only one class is left. The calculation of the classification accuracy at each level permits to determine the best number of classes for a desirable level of accuracy.

MONITORING VEGETATION DYNAMICS INFERRED BY SATELLITE DATA USING THE PHENOSAT TOOL

1. Summary

A long history of research in phenology has created a significant knowledge and ability to predict plant developmental events (McMaster, 2005).

During the last decade, the interest in phenology has been increasing significantly (e.g. Jonsson and Eklundh, 2004; Fisher et al., 2006; Hermance et al., 2007; Hudson et al., 2009). Phenology provides valuable information for many areas as land-use planning, crop management, carbon sequestration and more recently to infer about climate changes. The knowledge of timing of phenological events and their variability can help to get more stable crop yields and quality.

Focused on the crop management's importance in the actual general practices, a new software package (PhenoSat) was developed during this PhD project that allows the extraction of phenological information from EOS VI time-series.

Different land cover types (vineyard, semi-natural meadows, low shrublands), in different geographical locations of continental Portugal, were used to test the ability of PhenoSat in determining the phenological information. Two important features were also tested: the advantage of selecting an in-season temporal region of interest; and the capability to detect more than one growth cycle through a year.

Monitoring Vegetation Dynamics Inferred by Satellite Data Using the PhenoSat Tool

Arlete Rodrigues, *Student Member, IEEE*, Andre R. S. Marçal, and Mário Cunha

***Abstract*—PhenoSat is an experimental software tool that produces phenological information from satellite vegetation index time series. The main characteristics and functionalities of the PhenoSat tool are presented, and its performance is compared against observed measures and other available software applications. A multiyear experiment was carried out for different vegetation types: vineyard, low shrublands, and seminatural meadows. Temporal satellite normalized difference vegetation index (NDVI) data provided by MODerate resolution Imaging Spectroradiometer and Satellite Pour l'Observation de la Terre VEGETATION were used to test the ability of the software in extracting vegetation dynamics information. Three important PhenoSat features were analyzed: extraction of the main growing season information, estimation of double growth season parameters, and the advantage of selecting a temporal region of interest. Seven noise reduction filters were applied: cubic smoothing splines, polynomial curve fitting, Fourier series, Gaussian models, piecewise logistic, Savitzky–Golay (SG), and a combination of the last two. The results showed that PhenoSat is a useful tool to extract NDVI metrics related to vegetation dynamics, obtaining high significant correlations between observed and estimated parameters for most of the phenological stages and vegetation types studied. Using the combination of SG and piecewise logistic to fit the NDVI time series, PhenoSat obtained correlations higher than 0.71, except for the seminatural meadow start of season. The selection of a temporal region of interest improved the fitting process, consequently providing more reliable phenological information.**

***Index Terms*—Normalized difference vegetation index (NDVI), phenology, PhenoSat, time series.**

I. Introduction

TEMPORAL vegetation profiles based on remotely sensed data provide valuable information for understanding land cover dynamics, generally interpreted by vegetation phenological events. Over the last years, the number of remote sensing sensors capable of acquiring phenological appropriate temporal data has been increasing considerably. Most of these sensors [Advanced Very High Resolution Radiometer, Satellite Pour l'Observation de la Terre (SPOT) VEGETATION (SPOT VGT), MODerate resolution Imaging Spectroradiometer (MODIS), and MEdium Resolution Imaging Spectrometer] provide data related to a specific vegetation index (VI), such as the normalized difference vegetation index (NDVI) or the enhanced VI [1]. The VI temporal profile (time series) can be analyzed through time and used to extract information about the vegetation dynamics. Currently, access to satellite vegetation data is widespread, with low or no costs. However, the analysis and extraction of relevant information can be a difficult and time-consuming process due to the large amounts of data and the presence of noise.

The VI time series obtained by Earth Observation Satellite (EOS) images generally include various noise components such as atmospheric disturbances, viewing and solar illumination variability, cloud cover, and others. In the SPOT VGT (ten days) and MODIS (16 days) composites, some of this noise is reduced by the maximum value compositing process [2], where only the highest VI value in a predefined period (10 or 16 days according to the sensor) is retained. However, additional noise may be also introduced by the process of overlaying several images (for example, due to image registration). Noise reduction or fitting a model to the observed data is thus necessary before the extraction of vegetation dynamics information. It is important to carefully evaluate the time series and the noise present in order to choose the best fitting algorithm, one capable of smoothing the data without introducing artifacts or suppressing natural variations of vegetation [3]. Various time-series fitting algorithms have been developed [2], [4]–[10], and comparison studies have been conducted [11], which concluded that asymmetric Gaussian (AG) [6] and double logistic functions [4] present the best results.

During the last years, more fitting algorithms (e.g., [12]–[14]) and software products were developed to extract phenology from satellite VI time series. These software include TIMESAT [7], TimeStats [15], enhanced TIMESAT [16], PPET [17], and the software developed by United States Geological Survey Earth Resources Observation and Science Center [18]. All these software tools have important functionalities for the extraction of phenological information, but none of them allow the selection of an in-season window of interest, which is fundamental for application to crops with discontinuous canopy. Moreover, except for TIMESAT, none of them have a specific option to determine a double growth season phenology. To address these limitations, the authors of this paper developed PhenoSat, a new tool that is flexible to detect the number of growth seasons in each year and has an option that permits

Manuscript received February 29, 2012; revised July 12, 2012 and September 12, 2012; accepted September 30, 2012. Date of publication November 26, 2012; date of current version March 21, 2013. This work was supported by the Fundação para a Ciência e a Tecnologia under Doctoral Grant SFRH/BD/62189/2009.

A. Rodrigues and M. Cunha are with the Department of Geosciences, Environment and Spatial Planning and the Centro de Investigação em Ciências Geo-Espaciais, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal (e-mail: dr.arlete@gmail.com; mcunha@mail.icav.up.pt).

A. R. S. Marçal is with the Department of Mathematics and the Centro de Investigação em Ciências Geo-Espaciais, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal (e-mail: andre.marcal@fc.up.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2012.2223475

to define, manually or automatically, an in-season window of interest.

The aim of this paper is to evaluate the capability of NDVI metrics obtained from PhenoSat to identify the phenology *in situ*. Different NDVI time series provided by MODIS and SPOT VGT satellite sensors, from different geographical locations and vegetation types, were used to address the objectives of this paper. The following section describes PhenoSat functionalities and insights into the algorithm performance. Sections III and IV present an experiment carried out to test the PhenoSat ability to detect accurate phenology with various vegetation types, and Section V presents the concluding remarks.

II. PhenoSat Software Description

A. Functionalities and Implementation

PhenoSat is a simple-to-use tool that extracts satellite VI metrics related to vegetation phenology. PhenoSat was implemented in Matlab, receiving as input a temporal VI data set and as outputs the phenological information and the data from the various fitting steps.

PhenoSat extracts a number of phenological parameters for the main growth period such as the following: start of the season, maximum vegetation growth, or senescence. Some vegetation types can have a regrowth in the same year, which is related to some factors such as the grass-cut-animal pastures or particular weather conditions. The beginning and maximum of the regrowth period can be determined by PhenoSat.

Some VI data sets available online (e.g., MODIS and SPOT VGT) are already preprocessed in order to remove many of the disturbances provided by different factors such as atmospheric conditions or geometry and illumination variability [19]–[22]. Although this preprocessing is generally effective, the VI data sets still retain enough problems (punctual outliers or abrupt changes) that require additional processing. The elimination of these artifacts can be achieved by the application of noise reduction filters, which permits the researcher to conduct a better subsequent analysis and to obtain more reliable vegetation dynamics information.

Initially, the VI time-series values that are substantially different (with an NDVI difference above 0.2) from the left and right spatial neighbors, and from a median of a window (Mw), are considered outliers, and their values are replaced by the Mw value. PhenoSat has an option that permits the application of an upper envelope [4], enhancing the spring and summer periods. Despite these actions, some noise might still remain in the time series. Further improvements in the analysis can be obtained by PhenoSat using seven smoothing algorithms: cubic smoothing splines (CSSs), polynomial curve fitting (PCF), Gaussian models (GMs), Fourier series (FS), piecewise logistic (DL), Savitzky-Golay (SG), and, also, the sequential combination of the SG and DL.

The CSS algorithm fits a spline or smooth piecewise-polynomial algorithm, and a smoothing parameter determines just how closely the smoothing spline adheres to the given data [23].

The PCF [24] applies a polynomial of a given degree to fit the data. The higher the degree, the closer the fitting curve will be to the given data.

The FS [25] is a sum of sine and cosine functions of different period that describes a periodic signal. It is represented in either the trigonometric form or the exponential form

$$y = a_0 + \Sigma_{i=1}^{n} \left[a_i \cos(nwx) + b_i \sin(nwx)\right] \tag{1}$$

where a_0 models a constant (intercept) term in the data and is associated with the $i = 0$ cosine term, w is the fundamental frequency of the signal, and n is the number of terms (harmonics) in the series $(1 \leq n \leq 8)$.

The GM [26] fits peaks and is given by

$$y = \Sigma_{i=1}^{n} a_i e^{\left[-\left(\frac{x-b_i}{c_i}\right)^2\right]} \tag{2}$$

where a is the amplitude, b is the centroid (location), c is related to the peak width, and n is the number of peaks to fit $(1 \leq n \leq 8)$.

The SG is a particular type of low-pass filter, well adapted for data smoothing [27]. This filter

$$g_i = \Sigma_{n=nL}^{nR} c_n f_i + n \tag{3}$$

replaces each data value f_i, $i = 1, \ldots, N$, by a linear combination g_i of nearby values in a window defined by the number of points used "to the left" (nL) and "to the right" (nR) of a data point i, respectively. The application of the SG in PhenoSat uses the simplest case $(nL = nR)$, where the same number of points is used "to the left" and "to the right." Some tests were done to estimate the best SG moving window and polynomial order, and a first-degree polynomial with a frame size of five proved to be capable of removing the undesirable data from NDVI time series, capturing efficiently all the transitions related to the main and double growth seasons.

The DL, defined by (4), uses seven parameters to fit the VI data

$$\mathrm{VI}_t = \mathrm{VI}_w + \frac{k}{1+\exp\left[-c(t-p)\right]} - \frac{k+\mathrm{VI}_{w-\mathrm{VI}_{w1}}}{1+\exp\left[-d(t-e)\right]} \tag{4}$$

where t represents the time, VI_t is the VI value at time t, k is related with the VI asymptotical value, c and d are the slopes at "left" and "right," and p and e are the dates of the inflection points. VI_w and VI_{w1} are the VI values before the start of growing season and after the leaf fall, respectively. The continuity between the vegetation growth and senescence parts is assured by the k parameter, even when they differ in shape [28]. The parameters of the DL method were estimated by the Levenberg–Marquardt [29] algorithm that requires some reasonable initial values.

Fig. 1 presents a schematic representation of the DL parameters, using two consecutive years of NDVI SPOT VGT data. The continuity between the two years is assured by the NDVI_w and NDVI_{w1}, being the NDVI_w for the second year (beginning of the time series) the same as the NDVI_{w1} for the first year (final of the time series).

A new fitting method using a combination of SG and DL (SG+DL) was also implemented. It consists of the application

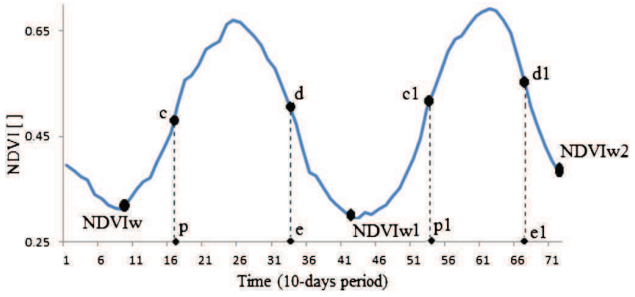


Fig. 1. Schematic representation of the DL parameters, using two consecutive years of NDVI SPOT VGT data.

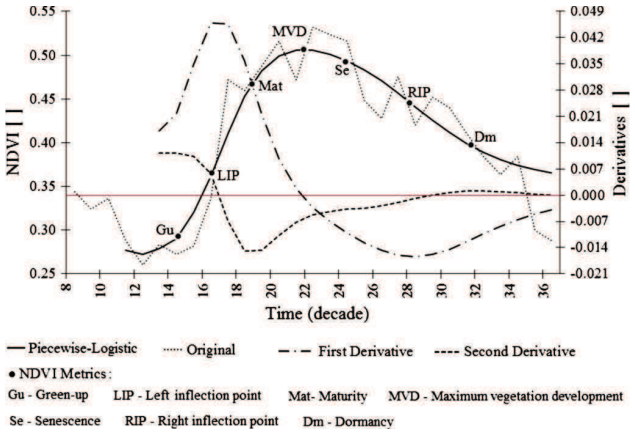


Fig. 2. Representation of NDVI metrics (Gu, LIP, Mat, MVD, Se, RIP, and Dm) calculated based on maximum and minimum rates of curvature change. Global maximum and minimum of the curvature estimate the inflection points.

of the SG filter to the NDVI time series, followed by the DL fitting process.

B. Phenological Information Extraction

PhenoSat extracts information (VI value and time of occurrence) of seven phenological stages in the main growing season: green-up (Gu) or start of season (SOS), left and right inflection points (LIP and RIP), maturity (Mat), maximum vegetation development (MVD), senescence (Se), and dormancy (Dm). To estimate these stages, the fitted data derivatives are used. Initially, the LIP and RIP, corresponding to the maximum growth and maximum senescence rates, are calculated using the maximum and minimum of the fitted data first derivative. Then, the MVD is determined as the maximum VI value observed in the data interval delimited by the inflection points. The Gu and Dm correspond to the maxima of the fitted data second derivative to the left and right of the MVD, respectively. Furthermore, the maturity [beginning of ripening stage (fruit vegetation) or full canopy (nonfruit vegetation)] and senescence correspond to the minima of the fitted second derivative to the left and right of the MVD. As an illustration, Fig. 2 shows the NDVI values optimized after DL application and a schematic procedure for the calculation of the NDVI metrics. The NDVI values were obtained from SPOT VGT sensor for a vineyard region in Portugal, for the year 2000.

In addition to the phenological information extracted by PhenoSat for the main growing season, it is also possible to record information for a double growth season or regrowth occurrence. This option allows obtaining information about the VI value and time occurrence for the start and maximum of this period. To determine the regrowth stage occurrence, the use of two successive years of data is necessary: the data for the year in analysis and the data for the following year. The start occurs when an increase of three or more consecutive points is verified after the Dm stage. The maximum of the double growth happens when, after the start occurrence, a decreasing period is verified.

Another feature of PhenoSat is the possibility of selecting an annual time-series subinterval, reducing the volume of data to be processed and thus improving the fitting process. This temporal region of interest can be selected manually, by the user, or automatically, by PhenoSat. The manual selection can be done by introducing the initial and final time positions in a specific application defined for this purpose. This selection must be based on the knowledge of the vegetation behavior, at normal growth conditions. For the automatic detection of the temporal region of interest, the PhenoSat first finds the maximum value of the data and then searches for the point where a significant increase (or abrupt decrease) is verified to the left of the maximum. That point corresponds to the initial position of the region of interest. Afterward, to determine the final position, the program proceeds in a similar way but evaluates the data to the right of the maximum.

III. Experimental Setup

A. Satellite Data and Study Areas

In order to evaluate the performance of PhenoSat, various tests were carried out using NDVI time series acquired from different sensors, representing a variety of vegetation types and geographical locations. Satellite vegetation data covering Portugal from Aqua MODIS NDVI (16-day composite and 250-m resolution) and NDVI SPOT VGT (ten-day composite and 1-km resolution) were downloaded from Reverb [30] and Vision on Technology (VITO) [31], respectively. Three different vegetation types, namely, vineyard (VIN), seminatural meadows (SNM), and low shrublands (LSL), were used for the experimental test carried out (Table I). These land cover types were selected due to the availability of VIN ground-based information as well as for its different growth patterns: 1) The grapevines in Portugal have a long dormancy period with intense understory vegetation growth and a discontinuous canopy [28]; 2) SNM presents a vegetation regrowth by August whose intensity and date of occurrence are mainly dependent on climatic conditions [32]; and 3) LSL, composed mainly of shrubs and permanent herbaceous species [33], is an extensive grazing area with minimal agronomic human intervention, where vegetation development occurs later than in the SNM.

For each test site selected, the median NDVI value was recorded from each image. The median was used instead of the mean values as it has lower sensitivity to erroneous or outlier values. Annual NDVI time series were created for VIN, SNM, and LSL.

TABLE I
DESCRIPTION OF STUDY AREAS AND SATELLITE DATA SETS USED

Land Cover Type	Code	Location	Coordinates (Lat/Long WGS84)	Satellite product	Time-series period	Size (pixels)
Vineyard (*)	VIN	Portugal- Douro region	UL: 7d25'09W, 41d15'21N BR: 7d23'37W, 41d13'53N	SPOT VGT (1km; 10 days)	2000 to 2005	9
Low Shrublands	LSL	Portugal- Montalegre	UL: 7d55'25W, 41d50'11N BR: 7d54'53W, 41d48'35N	SPOT VGT (1km; 10 days)	1999 to 2006	48
Semi-Natural Meadows	SNM	Portugal- Montalegre	UL: 7d57'19W, 41d37'58N BR: 7d56'58W, 41d37'34N	AQUA MODIS (250m; 16 days)	2003 to 2007	6

(*) Ground-based measurements available.

B. PhenoSat-Derived Phenology

For each test site, the SOS and MVD phenological stages were extracted. If the crop shows a double growth season, the start and maximum of this period were also recorded. The PhenoSat-derived phenology, for each vegetation type, was compared against observed measures and the results produced by TIMESAT. Both software tools extract phenological information from satellite VI time series. TIMESAT was chosen for this experiment because it is a freely available software tool and has time-series fitting algorithms similar to PhenoSat.

TIMESAT uses three distinct methods to fit the NDVI time series: AG, DL, and SG. Previous studies showed that AG and DL algorithms present similar results [11], [34]. As described in Section II-A, PhenoSat permits the use of seven different methods to smooth VI time series. However, for consistency, in this experiment, we only used DL, SG, and SG+DL as they are implemented in both software tools.

The SG filter was implemented in both tools in a similar way, replacing each data value by a linear combination of nearby values in a window [27]. The implementation of the DL differs in the number of parameters: PhenoSat uses seven parameters (see (4) in Section II-A), and TIMESAT uses only four parameters, which are related to the inflection points and the respective rates of change.

To determine the phenological metrics, TIMESAT uses a threshold approach: The start time of the season corresponds to the time for which the left edge has increased by 20% (default value used by TIMESAT) of seasonal amplitude, measured from the left minimum level. The time for the middle of the season is obtained by calculating the mean value of the times for which the left edge has increased to the 80% level and the right edge has decreased to the 80% level. On the other hand, PhenoSat uses an algorithm based on derivatives avoiding thresholds or empirical constants, providing a method that can be applied globally and is capable of identifying multiple growth periods within a single year.

C. Software Performance Evaluation

To evaluate the performance of the software in the extraction of phenological metrics, a Spearman's rank correlation (e.g., [35]) and a root-mean-square error (rmse) analysis were performed to compare the estimated and observed phenological measures. Table II presents the statistics of observed phenological measures obtained for each study area. For VIN, the observed measures were collected in the field, according to the Baggiolini scale [36]; for LSL and SNM, they were derived by the analysis of the original NDVI time series, taking into account the knowledge of the vegetation behavior in the field under normal conditions (regular atmospheric conditions, without human interventions that affect the normal growth of vegetation). The SOS was determined as the first point where a continuous (five or more points) vegetation growth was verified (March/April). The MVD was identified as the maximum NDVI value in the annual time series, which generally occurs in June or early July. The SNM regrowth was related to the first significant (three or more points) vegetation growth after the MVD (around August).

TABLE II
STATISTICS OF OBSERVED PHENOLOGICAL MEASURES OBTAINED FOR VINEYARD (VIN), LOW SHRUBLAND (LSL), AND SEMINATURAL MEADOW (SNM) VEGETATION TYPES. NOTE THAT THE OBSERVED MEASURES WERE OBTAINED FROM FIELD MEASUREMENTS (VIN) OR BY THE ANALYSIS OF THE NDVI ANNUAL PROFILES (LSL AND SNM)

Phenological Stage	Statistics	VIN	LSL	SNM
Start of Season (SOS)	Mean (DOY)	82.16	88.75	102.40
	Maximum (DOY)	91.00	100.00	128.00
	Minimum (DOY)	79.00	70.00	96.00
	Standard deviation (days)	4.44	9.91	14.31
Maximum Vegetation Development (MVD)	Mean (DOY)	206.66	177.50	150.40
	Maximum (DOY)	215.00	180.00	176.00
	Minimum (DOY)	201.00	170.00	128.00
	Standard deviation (days)	6.21	4.62	18.24

Note: The observed measures were obtained from field measurements (VIN) or by the analysis of the NDVI annual profiles (LSL and SNM).

Two other important points to evaluate are the ability of TIMESAT and PhenoSat to detect a double growth season and

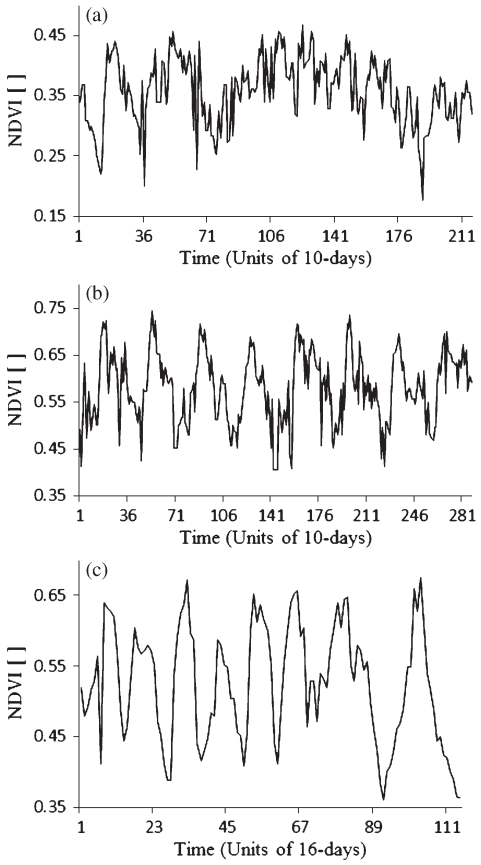


Fig. 3. Original NDVI temporal profiles for three land cover types: (a) Vineyard, (b) low shrubland, and (c) seminatural meadows. The x-axis scale of the plot (c) is different from the other two as the sensor used was MODIS (16 days), and for plots (a) and (b), the sensor used was SPOT VGT (ten days).

the advantage of the option of selecting a temporal region of interest (available in PhenoSat but not in TIMESAT).

IV. Results and Discussion

The original NDVI temporal profiles produced for each test site are shown in Fig. 3. As described, these plots were produced from the median NDVI of the test area available for each land cover type. The general shape of these temporal plots is very much what could be expected for the vegetation types. The SNMs are irrigated all year round, and in general, they present a SOS in March/April, an MVD in June/July, and a regrowth by August. LSL presents a similar behavior to SNM; however, as the LSLs are not irrigated, the regrowth is less frequent and with less magnitude.

For the VIN vegetation type, the annual profile presents high variability among years and is not so well defined as for SNM and LSL. Grapevine (VIN) budbreak occurs in March/April (Julian Day 79 to 91), followed by a period of about four weeks of intensive growth and then a steady decrease until "veraison" (change of color of the grapes) that occurs between Julian Day 201 and 215 (Table II; Fig. 3). Due to the large interrow space and discontinuous canopy, it is possible to see, particularly in the annual winter season, the effect of the soil vegetation growth on the VI temporal profiles.

A. PhenoSat Fitting Methods

To evaluate the ability of the PhenoSat fitting algorithms (SG, DL, SG+DL, CSS, PCF, GM, and FS) in estimating phenology, a comparison was made between derived and observed phenological information.

The performance of CSS, PCF, GM, and FS algorithms depends on a smoothing parameter, which affects the adherence of the fitting curve to the original data. To test the sensitivity of the algorithms to the smoothing parameter, PhenoSat was executed eight times for each algorithm: For PCF, GM, and FS, the parameter was varied between one and eight, and for CSS, it was varied from 0.1 to 0.8.

Table III presents the Spearman's correlation rank (rs) between PhenoSat-estimated phenological parameters and observed values, for the VIN region. The SOS and MVD obtained from PhenoSat were compared with grapevine field measures of budbreak and veraison ("change of color of the grapes"), respectively. The SG+DL algorithm presented correlations above 0.70 for both phenological stages evaluated. For CSS, FS, GM, and PCF algorithms, high smoothing parameters allowed a more accurate detection of the SOS, presenting correlations of 0.71, 0.83, 0.89, and 0.64, respectively. However, for the MVD, intermediate values obtained the best correlations except for CSS, where the lowest parameter (0.1) produced the highest correlation (0.40).

B. PhenoSat and TIMESAT Results

Table IV shows the Spearman's rank correlation (rs) and rmse results obtained between observed and estimated phenological SOS and MVD, for VIN, LSL, and SNM vegetation types.

Analyzing the fitting algorithms, SG+DL was the best for PhenoSat with correlations above 0.70 for all tests, except for the seminatural meadow SOS. In TIMESAT, SG and SG+DL were the algorithms that obtained the better and the worst correlations, respectively, between derived and observed phenologies.

For VIN, PhenoSat showed the best results in all fitting methods, providing rmse values between 15 and 23 days. In TIMESAT, some VIN years are not considered in the correlation estimation as the software was unable to predict the metrics for those years. The poor adherence of the TIMESAT fitting algorithms to the VIN original data can explain the abnormal rmse (above 100 days) results obtained.

For all vegetation types, TIMESAT and PhenoSat proved to be useful tools to detect phenological events, obtaining, in some cases, correlations above 0.70. TIMESAT obtained better results for the MVD except using the SG+DL algorithm, for which the PhenoSat always was the best with mean correlations of about 0.70 and rmse values between 5 and 23 days.

Table V presents the observed and computed regrowth season parameters. In 2007, there was no regrowth, which was correctly verified by both software tools. For the other years, TIMESAT had some difficulty in determining the double growth season using the different algorithms; the SG+DL was the worst case, as TIMESAT was unable to determine the parameters for any year tested. The TIMESAT 1-season option does not report any regrowth; thus, the results presented in

TABLE III
SPEARMAN'S CORRELATION RANK (rs) BETWEEN PHENOSAT-ESTIMATED PHENOLOGICAL PARAMETERS AND OBSERVED MEASURES, FOR THE VINEYARD TEST SITE

Algorithm	Start of Season (SOS)			Maximum Vegetation Development (MVD)		
	Best Parameter	*rs*	*p-value*	*Best Parameter*	*rs*	*p-value*
DL	—	-0.89	0.009	—	-0.09	0.44
SG	—	-0.15	0.39	—	0.15	0.39
SG+DL	—	-0.77	0.04	—	-0.71	0.06
CSS	0.8	-0.71	0.06	0.1	-0.40	0.22
PCF	8	-0.64	0.09	4	0.62	0.10
FS	8	-0.83	0.02	6	-0.43	0.20
GM	8	-0.89	0.01	5	-0.60	0.10

DL, SG , SG+DL, CSS, PCF, FS and GM represent the 7 fitting algorithms, respectively, piecewise-logistic, Savitzky-Golay, combination of Savitzky-Golay and piecewise-logistic, cubic smoothing splines, polynomial curve fitting, Fourier series and Gaussian models.

TABLE IV
SPEARMAN'S RANK CORRELATION (rs) AND RMSE BETWEEN OBSERVED AND ESTIMATED PHENOLOGIES, OBTAINED FROM TIMESAT AND PHENOSAT, FOR VINEYARD (VIN), LOW SHRUBLAND (LSL), AND SEMINATURAL MEADOW (SNM) VEGETATION TYPES

Statistics	DL				SG				SG+DL			
	TIMESAT		PhenoSat		TIMESAT		PhenoSat		TIMESAT		PhenoSat	
	SOS	*MVD*	*SOS*	*MVD*	*SOS*	*MVD*	*SOS*	*MVD*	*SOS*	*MVD*	*SOS*	*MVD*
VIN												
n	5	5	6	6	4	4	6	6	3	3	6	6
rs	0.10	-0.60	-0.89	-0.09	-0.40	-0.80	-0.15	0.15	-0.50	--	-0.77	-0.71
p-value	0.44	0.14	0.009	0.44	0.30	0.10	0.39	0.39	0.33	--	0.04	0.06
RMSE	*	*	18	23	*	*	15	19	*	*	15	23
LSL												
n	8	8	8	8	8	8	8	8	8	8	8	8
rs	0.43	-0.13	0.56	-0.13	0.55	0.76	0.43	0.61	0.12	0.38	0.79	0.76
p-value	0.14	0.38	0.07	0.38	0.08	0.02	0.14	0.06	0.39	0.18	0.01	0.02
RMSE	*	11	9	17	11	11	13	8	11	10	5	5
SNM												
n	5	5	5	5	5	5	5	5	5	5	5	5
rs	-0.71	0.87	0.71	0.05	-0.35	0.98	-0.41	0.92	-0.35	0.41	0.35	0.95
p-value	0.09	0.03	0.09	0.47	0.28	0.002	0.25	0.01	0.28	0.27	0.28	0.007
RMSE	8	11	10	19	10	14	7	7	10	28	17	10

SOS represents the start of season and MVD the maximum vegetation development.
rs is the Spearman's rank correlation coeficient and p-value the probability associated. n is the number of observations.
DL, SG and SG+DL represent the 3 fitting algorithms, respectively, piecewise-logistic, Savitzky-Golay and the combination of Savitzky-Golay with piecewise-logistic.
RMSE is the Root Mean of Square Error and * represents a RMSE value above 100 days.

Table V were obtained using the option 2-seasons. Although this option might work for some years, it is not sufficiently flexible for the years in which there is no regrowth. On the other hand, PhenoSat determined the parameters with high precision, only failing on the maximum for the year 2005. The PhenoSat algorithm, to derive the regrowth, uses the preprocessed time series before the application of the fitting process, being independent from the fitted data, and thus provides identical results for all fitting algorithms tested.

Fig. 4 shows the original and SG, DL, and SG+DL fitted data, obtained by PhenoSat and TIMESAT software tools, for the SNM region in 2003. From the original data, it is possible to verify a regrowth occurrence around the Julian Day 240 to 256. The TIMESAT SG algorithm cannot detect efficiently the regrowth, smoothing this period, contrary to PhenoSat that can detect with high precision the beginning of the regrowth. For DL, the TIMESAT and PhenoSat fitted results present slight differences for the main growing season; however, for the regrowth period, PhenoSat tends to smooth this period. The SG+DL algorithm presents a similar performance to that verified in SG for the TIMESAT tool; for PhenoSat, the SG+DL fitted results show a "mix" of the SG and DL results.

TABLE V
OBSERVED AND ESTIMATED DOUBLE GROWTH SEASON PARAMETERS, FROM TIMESAT (2-SEASONS) AND PHENOSAT, FOR A SEMINATURAL MEADOW REGION

Year	DL				SG				SG+DL				OBSERVED	
	TIMESAT		PhenoSat		TIMESAT		PhenoSat		TIMESAT		PhenoSat			
	Start	*Max*	*Start*	*Max*	*Start*	*Max*	*Start*	*Max*	*Start*	*Max*	*Start*	*Max*	*Start*	*Max*
2003	238	291	252	304	290	333	252	304	n.d.	n.d.	252	304	252	304
2004	229	283	252	288	n.d.	n.d.	252	288	n.d.	n.d.	252	288	252	288
2005	262	321	252	336	264	326	252	336	n.d.	n.d.	252	336	252	328
2006	235	305	252	320	244	300	252	320	n.d.	n.d.	252	320	252	320
2007	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

n.d. signifies that no regrowth is verified on this year.
DL, SG and SG+DL represent the 3 fitting algorithms, respectively, piecewise-logistic, Savitzky-Golay and the combination of Savitzky-Golay with piecewise- logistic.

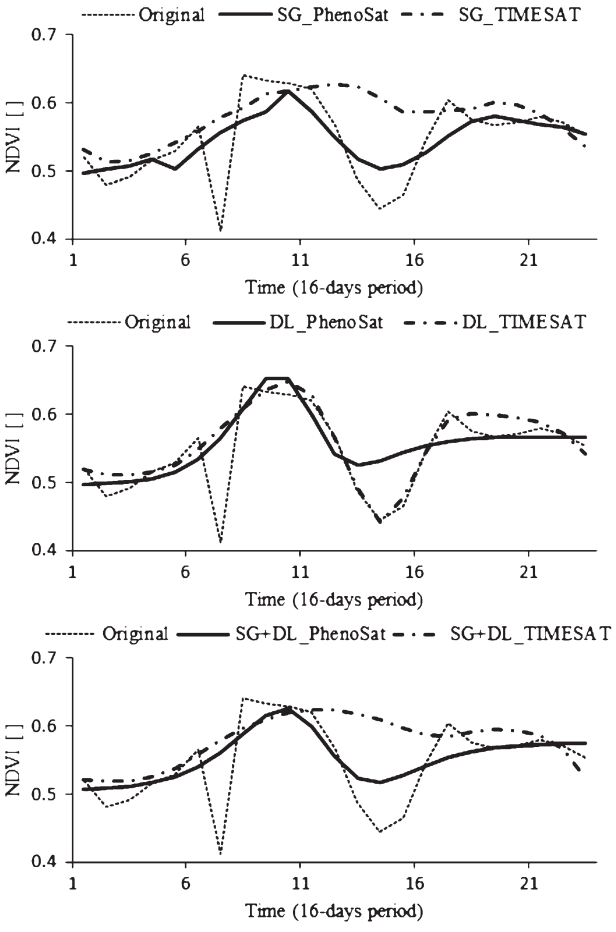


Fig. 4. Original and fitted results obtained from PhenoSat and TIMESAT using SG, piecewise logistic (DL), or both (SG+DL) algorithms, for the seminatural meadow vegetation type in 2003.

The selection of an in-season window is of particular interest for crops like VIN, with a long period of winter dormancy, and crops with discontinuous canopies. The VIN, in general, presents a budbreak in March/April (Julian Day 79 to 91) and dormancy stage in October/November (Julian Day 300 to 310). Thus, the period comprising the Julian Day 79 to 310 is usually used to define a manual selection of the region of interest. However, the manual selection, based on the knowledge of the vegetation behavior in the field under normal conditions, is not flexible to adapt to the different vegetation dynamics that can occur over the years due to some factors, such as adverse weather conditions. The automatic selection can be a good option to eliminate these limitations and detect more accurately the temporal region of interest.

Fig. 5 shows the original VIN data for 2000 and 2003 and SG+DL fitted data obtained using all range of observations and using a region of interest automatically selected by PhenoSat. For the year 2000, PhenoSat selects the range Julian Day 110 to 310 as the temporal region of interest, which is easily recognized as the VIN dynamics period in the original data. However, for the year 2003, only 15 observations (Julian Day 100 to 240) are selected. The difference in the two years can be explained by the fact that, in 2003, there was an irregular precipitation from August 30 to September 5, which was not verified in 2000. The weather conditions in 2003 caused an early growth of the soil vegetation, which can be seen graphically around the Julian Day 250.

Although, for the year 2000, no advantage was verified in selecting a subtemporal region of interest, this is not verified in 2003, where the fitted data provided by all 36 observations tend to oversmooth the original data and are therefore unable to detect the regrowth occurrence around Julian Day 250. This difference in the fitted results leads to different phenological stage occurrences. Using all 36 observations ($r^2 = 0.41$), the SOS and the MVD were obtained on Julian Day 140 and Julian Day 160, respectively; however, using the 15 observations ($r^2 = 0.96$) provided by the region of interest, an earlier SOS (Julian Day 120) and a later MVD (Julian Day 180) were verified, these results being more consistent with the original data. This demonstrates that the selection of the temporal region of interest produces an improvement in the fitted process, thus obtaining more reliable results.

Table VI presents the effect of the temporal region of interest selection on the Spearman's correlations between observed and PhenoSat-derived phenologies, for the VIN region (2000–2005). Using the temporal region of interest, the correlations range between -0.70 and -0.77 being always higher than

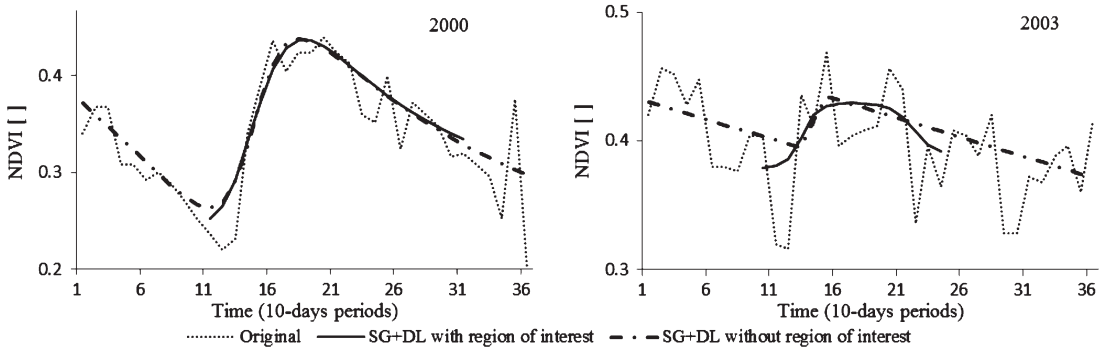


Fig. 5. (Dotted line) Original vineyard data for 2000 and 2003, (solid line) SG+DL fitted data obtained using a subtemporal region of interest, and (dashed line) all range of observations.

TABLE VI
CORRELATIONS (rs) BETWEEN OBSERVED AND PHENOSAT-DERIVED PHENOLOGY PARAMETERS FOR VINEYARD (2000–2005)

Phenological stages	In-season selected region	
	Not selected	Selected
Start of Season (n=6)	-0.51	-0.77
Flowering (n=6)	0.26	-0.70
Maximum Vegetation Development (n=6)	0.47	-0.71

Significant level of the coefficient correlation for 6 observations: rs>0.676 (0.10); rs>0.759 (0.05).

those obtained using the complete range of 36 observations. Thus, the PhenoSat feature to select an in-season region of interest proved to be a valuable tool for vineyard monitoring, allowing for high significant correlations between estimated and observed values to be obtained. Adverse weather conditions such as unseasonal snow, extreme heat, or irregular precipitation could result in a false vegetation regrowth. The selection of a subtemporal region of interest based on vegetation dynamics knowledge could help to deal with the false report of regrowth in particular natural environments (high latitudes or boreal regions), leading to a better analysis and more consistent results.

V. CONCLUSION

The experiment carried out to evaluate the performance of the PhenoSat showed that it can produce accurate and consistent results, compared with observed measures. This tool proved capable of solving some limitations present in other software tools, such as the following: the detection of a double growth season, with the extraction of phenological parameters for this period, and the possibility to select, manually or automatically, an in-season region of interest.

The extraction of phenological parameters using an algorithm based on derivatives allows PhenoSat to avoid thresholds or empirical constants, providing a method that can be applied globally and that is capable of identifying multiple crops or regrowth within a single year. The tool proved to be very efficient in detecting the double growth period. The independency of the fitted results leads to a more realistic time-series profile over the year and, thus, more accurate regrowth-derived results.

The option to select, automatically or manually, a temporal region of interest of the VI time series provides an improvement in the fitting process, leading to more reliable results. The PhenoSat feature to select an in-season region of interest proved to be a valuable tool for vineyard monitoring and can enlarge the PhenoSat application to crops with discontinuous canopy, like forestry and deciduous fruit trees.

PhenoSat is a freely available software tool, with a preliminary version currently available at http://www.fc.up.pt/LamSat_XXI. This version is able to run a single data set file (with data for a single vegetation type, for different years) at a time. However, a new version is being developed to improve and automate the process in order to receive various data set files at once. Another improvement consists in creating an application that permits the extraction of phenological information directly from the EOS images. This option eliminates the laborious and manual preprocessing steps that involve the extraction of VI data from the images to create the data set input file.

ACKNOWLEDGMENT

The authors would like to thank VITO for the SPOT VEGETATION and Reverb for the MODIS images and also to the authors of TIMESAT for providing access to their software tool.

REFERENCES

[1] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 195–213, Nov. 2002.

[2] B. N. Holben, "Characteristics of maximum-value composite images from temporal AVHRR data," *Int. J. Remote Sens.*, vol. 7, no. 11, pp. 1417–1434, Nov. 1986.

[3] F. Fontana, C. Rixen, T. Jonas, G. Aberegg, and S. Wunderle, "Alpine grassland phenology as seen in AVHRR, VEGETATION and MODIS NDVI time series—A comparison with *in situ* measurements," *Sensors*, vol. 8, no. 4, pp. 2833–2853, Apr. 2008.

[4] P. S. A. Beck, C. Atzberger, K. A. Hogda, B. Johansen, and A. K. Skidmore, "Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 321–334, Feb. 15, 2006.

[5] B. A. Bradley, R. W. Jacob, J. F. Hermance, and J. F. Mustard, "A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data," *Remote Sens. Environ.*, vol. 106, no. 2, pp. 137–145, Jan. 30, 2007.

[6] P. Jonsson and L. Eklundh, "Seasonality extraction by function fitting to time-series of satellite sensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1824–1832, Aug. 2002.
[7] P. Jonsson and L. Eklundh, "TIMESAT—A program for analyzing time-series of satellite sensor data," *Comput. Geosci.*, vol. 30, no. 8, pp. 833–845, Oct. 2004.
[8] M. G. Ma and F. Veroustraete, "Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China," *Adv. Space Res.*, vol. 37, no. 4, pp. 835–840, 2006, Natural Hazards and Oceanographic Processes from Satellite Data.
[9] A. Moody and D. M. Johnson, "Land-surface phenologies from AVHRR using the discrete Fourier transform," *Remote Sens. Environ.*, vol. 75, no. 3, pp. 305–323, Mar. 2001.
[10] G. J. Roerink, M. Menenti, and W. Verhoef, "Reconstructing cloudfree NDVI composites using Fourier analysis of time series," *Int. J. Remote Sens.*, vol. 21, no. 9, pp. 1911–1917, Jun. 15, 2000.
[11] J. N. Hird and G. J. McDermid, "Noise reduction of NDVI time series: An empirical comparison of selected techniques," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 248–258, Jan. 15, 2009.
[12] W. Zhu, Y. Pan, H. He, L. Wang, M. Mou, and J. Liu, "A changing-weight filter method for reconstructing a high-quality NDVI time series to preserve the integrity of vegetation phenology," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1085–1094, Apr. 2012.
[13] H. Carrao, P. Goncalves, and M. Caetano, "A nonlinear harmonic model for fitting satellite image time series: Analysis and prediction of land cover dynamics," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1919–1930, Apr. 2010.
[14] J. M. Chen, F. Deng, and M. Z. Chen, "Locally adjusted cubic-spline capping for reconstructing seasonal trajectories of a satellite-derived surface parameter," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2230–2238, Aug. 2006.
[15] T. Udelhoven, "TimeStats: A software tool for the retrieval of temporal patterns from global satellite archives," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 2, pp. 310–317, Jun. 2011.
[16] B. Tan, J. T. Morisette, R. E. Wolfe, F. Gao, G. A. Ederer, J. Nightingale, and J. A. Pedelty, "An enhanced TIMESAT algorithm for estimating vegetation phenology metrics from MODIS data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 2, pp. 361–371, Jun. 2011.
[17] R. D. McKellip, K. W. Ross, J. P. Spruce, J. C. Smoot, R. E. Ryan, G. E. Gasser, D. L. Prados, and R. D. Vaughan, "Phenological parameters estimation tool," NASA Tech. Briefs, New York, Sep. 30, 2010.
[18] K. W. Ross, B. A. Spiering, and M. T. Kalcic, "Monitoring phenology as indicator for timing of nutrient inputs in northern gulf watersheds," in *Proc. OCEANS*, 2009, pp. 1–4, vols 1–3.
[19] J. M. B. Carreiras, J. M. C. Pereira, Y. E. Shimabukuro, and D. Stroppiana, "Evaluation of compositing algorithms over the Brazilian Amazon using SPOT-4 VEGETATION data," *Int. J. Remote Sens.*, vol. 24, no. 17, pp. 3427–3440, Sep. 2003.
[20] B. Holben and R. S. Fraser, "Red and near-infrared sensor response to off-nadir viewing," *Int. J. Remote Sens.*, vol. 5, no. 1, pp. 145–160, Feb. 1984.
[21] G. G. Gutman, "Vegetation indexes from AVHRR—An update and future-prospects," *Remote Sens. Environ.*, vol. 35, no. 2/3, pp. 121–136, Feb./Mar. 1991.
[22] X. W. Li and A. H. Strahler, "Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy—Effect of crown shape and mutual shadowing," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 276–292, Mar. 1992.
[23] C. H. Reinsch, "Smoothing by spline functions," *Numerische Math.*, vol. 10, no. 3, pp. 177–183, 1967.
[24] J. Verschelde, "Introduction to symbolic computation: A maple/MATLAB course," in *Proc. Maple Conf.*, Waterloo, ON, Canada, 2005, pp. 500–509.
[25] *Curve Fitting Toolbox: Fit Curves and Surfaces to Data Using Regression, Interpolation, and Smoothing*, MathWorks, Natick, MA, 2011.
[26] A. Goshtasby and W. D. Oneill, "Curve-fitting by a sum of Gaussians," *CVGIP—Graph. Models Image Process.*, vol. 56, no. 4, pp. 281–288, Jul. 1994.
[27] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[28] M. Cunha, A. R. S. Marçal, and A. Rodrigues, "A comparative study of satellite and ground-based vineyard phenology," in *Proc. 29th Symp. EARSeL*, Chania, Greece, 2010, pp. 68–77.
[29] D. Montgomery, E. Peck, and G. Vining, *Introduction to Linear Regression Analysis*, 4th ed. Hoboken, NJ: Wiley, 2006.
[30] Reverb—The Next Generation Earth Science Discovery Tool, Jan. 13, 2012. [Online]. Available: https://reverb.echo.nasa.gov
[31] Vision on Technology, Jan. 13, 2012. [Online]. Available: http://free.vgt.vito.be/
[32] I. Pocas, M. Cunha, and L. S. Pereira, "Dynamics of mountain semi-natural grassland meadows inferred from SPOT-VEGETATION and field spectroradiometer data," *Int. J. Remote Sens.*, vol. 33, no. 14, pp. 4334–4355, Jul. 2012.
[33] I. Pocas, M. Cunha, and L. S. Pereira, "Remote sensing based indicators of changes in a mountain rural landscape of Northeast Portugal," *Appl. Geograph.*, vol. 31, no. 3, pp. 871–880, Jul. 2011.
[34] F. Gao, J. T. Morisette, R. E. Wolfe, G. Ederer, J. Pedelty, E. Masuoka, R. Myneni, B. Tan, and J. Nightingale, "An algorithm to produce temporally and spatially continuous MODIS-LAI time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 60–64, Jan. 2008.
[35] J. L. Myers and A. D. Well, *Research Design and Statistical Analysis*, 2nd ed. Mahwah, New Jersey: Lawrence Erlbaum Assoc. Inc., 2003.
[36] M. Baggiolini, "Les stades repères dans le développement annuel de la vigne et leur utilisation pratique," *Rev. romande Agric. Vitic. Arboric.*, vol. 8, pp. 4–6, 1952.

Arlete Rodrigues (S'12) was born in Avelar, Leiria, Portugal. She received the B.Sc. degree in mathematics from the Universidade de Aveiro, Aveiro, Portugal, in 2004 and the M.Sc. degree in remote sensing from the Faculdade de Ciências, Universidade do Porto, Porto, Portugal, in 2008, where she is currently working toward the Ph.D. degree.

Since 2007, she has been a Researcher with the Centro de Investigação em Ciências Geo-Espaciais (CICGE), Faculdade de Ciências, Universidade do Porto, with a research topic of temporal remote sensing data analysis for vegetation studies. Her scientific interests include image processing, pattern recognition, remote sensing, and time-series analysis.

Andre R. S. Marçal was born in Luanda, Angola. He received the B.S. degree in physics (solid-state physics) from the Universidade do Porto, Porto, Portugal, in 1991 and the M.Sc. and Ph.D. degrees in remote sensing and image processing from the University of Dundee, Dundee, U.K., in 1994 and 1998, respectively.

He is currently an Assistant Professor with the Department of Mathematics, Universidade do Porto. His research interests include various topics in remote sensing and image processing.

Prof. Marçal was a recipient of the prize for the Best Ph.D. from the Remote Sensing Society (U.K.) in 1999. From 2005 to 2010, he was a bureau member of the European Association of Remote Sensing Laboratories, as the Secretary General and later as the Vice-Chairman.

Mário Cunha was born in Barcelos, Porto, Portugal. He received the B.S. degree in agronomic engineering from the Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal, and the Ph.D. degree in agrarian sciences from the Faculdade de Ciências, Universidade do Porto (FCUP), Porto, Portugal.

He is currently an Assistant Professor with the Department of Geosciences, Environment and Spatial Planning, FCUP, where he is also a Researcher with the Remote Sensing, Image Processing and GIS Group, Centro de Investigação em Ciências Geo-Espaciais.

IDENTIFICATION OF POTENTIAL LAND COVER CHANGES ON A CONTINENTAL SCALE USING NDVI TIME-SERIES FROM SPOT VEGETATION

1. Summary

Land cover maps of global or continental scales are basic information for many kinds of applications (global change research, resource management). Accurate, timely information on the distribution of vegetation on the Earth's surface is a requisite for understanding how changes in land cover affect phenomena as diverse as the terrestrial primary productivity, the hydrological cycle and the energy balance at the surface-atmosphere interface.

The validation of land cover maps covering small areas is usually achieved by the use of ground data. However, for large areas this data collection can be problematic, due to the extent of the area concerned, the diverse ecosystems, the lack of homogeneous national legends, and the potential financial costs (Mayaux et al., 2008). Therefore, ground data are often replaced by EOS data (Strahler et al., 2006).

To update land cover maps for large scales, and avoid the logistical and financial problems in obtaining ground data, a new methodology was developed in this work. The methodology is based exclusively on satellite vegetation time-series and consists on a hierarchical aggregation of the land cover types, which is determined using the similarity between the yearly VI profiles.

The second largest continent of the world (Africa) is used to apply and evaluate the proposed approach. According to a statement by the United Nations Environment Programme (UNEP), the African Continent is particularly vulnerable to climate changes, and is suffering from deforestation at twice the average world rate. These characteristics reinforce the need for land use / land cover information to better understand the continent and the interactions between climate, ecosystem and human activities.

International Journal of Remote Sensing, 2013
Vol. 34, No. 22, 8028–8050, http://dx.doi.org/10.1080/01431161.2013.828184

Taylor & Francis
Taylor & Francis Group

Identification of potential land cover changes on a continental scale using NDVI time-series from SPOT VEGETATION

Arlete Rodrigues*, André R.S. Marçal, and Mário Cunha

Centro de Investigação em Ciências Geo-Espaciais, Universidade do Porto, Porto 4169–007, Portugal

(*Received 25 October 2012; accepted 12 July 2013*)

The identification of land cover changes on a continental scale is a laborious and time-consuming process. A new methodology is proposed based exclusively on SPOT VGT data, illustrated for the African Continent using GLC2000 as reference to select 26 distinct land cover types (classes). For each class, the normalized difference vegetation index (NDVI) time-series are extracted from SPOT VGT images and a hierarchical aggregation is done using two different methods: one that preserves the initial signatures throughout the hierarchical process, and another that recalculates the signatures for each aggregation level. The average classification agreement was above 89% using 26 classes. Reducing the number of classes improves classification agreement. In order to study the influence of temporal variability in the classification results, the methodology was applied on data from 1999, 2001, 2008, and 2010. With 26 classes, the best average classification agreement obtained was 94.5% with annual data, against 74.1% with interannual data.

1. Introduction

Land use and land cover maps are important to predict and manage the resources of the land surface, which can be related to economic, social, and environmental sustainability aspects. Studies based on land cover maps are usually carried out using remote-sensing data from either aerial surveys or Earth Observation Satellites (EOS). In the case of EOS, high-resolution images acquired by Landsat Thematic Mapper (TM) and Satellite Pour l'Observation de la Terre (SPOT) High Resolution Visible (HRV) have been widely used in the creation of a large number of land cover maps (e.g. Marçal et al. 2005; Pôças et al. 2011). These sensors have provided high-spatial resolution multispectral data since 1982 and 1986, respectively. Improvement in sensor characteristics over the years has opened up new possibilities for the production of accurate land cover maps due to the very high spatial (e.g. IKONOS satellite) or spectral resolution (e.g. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)).

Furthermore, the availability of daily (or 2–3 day periods) global coverage at moderate spatial resolution (300–1000 m pixels) provides high-temporal resolution data sets that can be valuable for many applications, including land cover change detection (e.g. Kleynhans et al. 2011; Lunetta et al. 2006; Zhan et al. 2002). The accuracy of the classification performed for land cover map production depends not only on the data preprocessing stages,

*Corresponding author. Email: dr.arlete@gmail.com

the classification methods, and the number of classes used, but also on the availability of training/control sample data. Various studies have proposed methods and algorithms for land cover classification at regional (e.g. Bosard, Feranec, and Otahel 2000; Homer et al. 2004; Vogelmann et al. 2001) and global scales (e.g. Bartholomé and Belward 2005; DeFries and Townshend 1994; DeFries et al. 1998; Friedl et al. 2002; Hansen et al. 2000; Loveland et al. 2000).

The automatic classification of satellite image data can be performed by two different strategies: multispectral images or temporal image sequences. Multispectral image classification focuses on finding patterns on the spectral response while taking into account known land cover groups. This approach usually uses high-spatial resolution images (single or several from different seasons) with multiple spectral bands. The temporal image classification approach uses image sequences acquired over a period of time, thus providing high-frequency observation of the land surface. However, these data sets can only realistically be obtained at low spatial resolution (typically 1 km pixel size). The data processing and distribution of these data sets have considerably improved in recent years, enabling access to high-level data (calibrated geometrically and corrected) at reduced or zero cost.

A number of classification experiments using temporal satellite data have been reported for different geographical areas, as summarized in Table 1. It is difficult to compare the classification accuracy of these studies because they use different numbers of classes (5–14) and various satellite data (Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Very High Resolution Radiometer (AVHRR), and SPOT

Table 1. Summary of classification results using EOS temporal data reported in various publications.

Study area	Satellite data	Classes	Best accuracy (%)	Author(s)
Canada	AVHRR 1 km	12	61	Latifovic and Pouliot (2005)
Sub-Saharan Africa	16 day MODIS 1 km	13	76	Vanacker et al. (2005)
China	16 day MODIS 500 m	9	88	Bagan, Wang, and Yasuoka (2007)
Portugal	8 day MODIS 500 m	9	90	Carrão, Gonçalves, and Caetano (2007)
South Africa	16 day MODIS 500 m	14	79	Colditz (2007)
Germany	16 day MODIS 500 m	14	83	Colditz (2007)
Volta Basin	16 day MODIS 250 m	10	77	Machwitz et al. (2008)
Kansas	8 day MODIS 250 m	6	94	Wardlow and Egbert (2008)
Dry Chaco Ecoregion	16 day MODIS 250 m	7	79	Clark et al. (2010)
Western Burkina Faso and Southern Mali	16 day MODIS 250 m	5	67	Landmann et al. (2010)
Krishna River Basin	8 day MODIS 250 m	9	85	Gumma, Thenkabail, and Nelson (2011)
Belgium	Monthly SPOT VGT 1 km/Monthly MODIS 250 m	8	80	Heremans et al. (2011)
North-Western Morocco	16 day MODIS 250 m	5	80	Hopfner and Scherer (2011)
Mali, West Africa	16 day MODIS 250 m	13	70	Vintrou et al. (2012)

VEGETATION (VGT)). Nevertheless, the accuracy values reported range between 61% with 12 classes (Latifovic and Pouliot 2005) and 94% with 6 classes (Wardlow and Egbert 2008).

Recent improvements in temporal data availability have made it possible to extract data for land cover map production at continental and global scales. However, the classification at these scales can be a difficult and time-consuming process due to limited available ground truth data (Loveland et al. 2000).

One of the land cover maps available for the African Continent is the Global Land Cover Map for the Year 2000 (GLC2000). GLC2000 was created under the coordination of the Joint Research Centre of the European Commission (Bartholomé and Belward 2005), making use of four types of remotely sensed data provided by SPOT VGT, the European Remote-Sensing Satellite (ERS), the Japanese Earth Resources Satellite (JERS), and the Defense Meteorological Satellite Program (DMSP). In addition, ancillary data from different sources were used as well as the expertise of 12 partners from seven local organizations (Mayaux et al. 2004). The diversity of data input, and the labour required for processing remote-sensing data and collection of ground truth data needed to produce GLC2000 are the main difficulties in frequent updates of this land cover map.

This article proposes a methodology to identify areas of potential change on a continental scale, with an application to the GLC2000 of Africa. The method is based exclusively on normalized difference vegetation index (NDVI) time-series from SPOT VGT, reducing the data input volume and time needed to produce updated land cover maps.

The potentially altered areas of GLC2000 are determined using a classification procedure based on the hierarchical aggregation of land cover classes. Initially, the GLC2000 data set is used to select 26 classes in the NDVI SPOT VGT images. The NDVI profiles are extracted for each class using the year 2000 as reference. A hierarchical structuring of classes is determined, which consists of the aggregation of the two most similar, level-by-level, until only one class is left. At each level, a similarity measure is applied and the two most similar classes are merged into a new class. Two distinct methods are used: one that preserves the initial profiles (signatures) in all hierarchical levels, and a second that uses the mean of the original signatures to create the signature of the new one. A cross-validation process (Geisser 1975; Stone 1974) is used to estimate the classification agreement for each class at each level of aggregation. The acquisition of ground truth data for classification assessment of large areas and temporal continuity is an uncommon practice due to limitations such as costs, time, and logistical challenges (Loveland et al. 2000). Due to the impossibility of acquiring ground truth data for this work, the term ‘classification agreement’ used in this article is related to GLC2000, which is considered as the reference.

In order to study the influence of temporal variability in the classification results, the methodology proposed was applied for the years 1999, 2001, 2008, and 2010. Two approaches were followed: one that considers the NDVI profiles of land cover used for training from GLC2000 unchanged over the years, using training from the year under analysis; and a second that eliminates the problems related to annual coverage, using training from the reference year (2000) and control samples from the test year (1999, 2001, 2008, or 2010).

2. Material and methods

2.1. Study area

The experimental work was done using image data from the African Continent, covering an area of 30.3 million km^2, from latitudes 37° N to 34° S and longitudes 18° W to 51° E.

Table 2. List of land cover classes used for Africa.

Class	Description	Code	No. pixels	Geographical centre
1	Bare Rock	BR	4800	20° 53′ N, 17° 49′ E
2	Croplands (>50%)	C	1702	27° 13′ S, 29° 07′ E
3	Closed Deciduous Forest	CDF	2408	12° 39′ S, 27° ° 00′ E
4	Closed Evergreen Lowland Forest	CELF	3528	3° 39′ N, 14° 11′ E
5	Closed Grassland	CG	3752	21° 11′ S, 46° 15′ E
6	Cities	Ct	576	26° 09′ S, 28° 07′ E
7	Croplands with Open Woody Vegetation	COWV	2170	12° 07′ N, 20° 04′ E
8	Degraded Evergreen Lowland Forest	DELF	162	24° 04′ S, 47° 20′ E
9	Deciduous Shrubland with Sparse Trees	DSST	1740	12° 00′ N, 10° 01′ W
10	Deciduous Woodland	DW	5715	7° 43′ N, 24° 49′ E
11	Irrigated Croplands	IC	2800	31° 03′ N, 30° 49′ E
12	Mangrove	M	139	11° 15′ N, 15° 19′ W
13	Mosaic Forest/Croplands	MFC	3822	6° 13′ N, 6° 28′ W
14	Mosaic Forest/Savanna	MFS	2668	4° 55′ N, 28° 37′ E
15	Montane Forest (>1500 m)	MF	1116	7° 49′ N, 35° 39′ E
16	Open Deciduous Shrubland	ODS	181	4° 20′ S, 13° 14′ E
17	Open Grassland	OG	2988	25° 22′ S, 22° 23′ E
18	Open Grassland with Sparse Shrubs	OGSS	2992	20° 12′ S, 19° 25′ E
19	Swamp Bushland and Grassland	SBG	1575	7° 59′ N, 30° 28′ E
20	Stony Desert	SD	7936	29° 24′ N, 11° 48′ E
21	Sandy Desert and Dunes	SDD	31373	19° 08′ N, 6° 35′ W
22	Submontane Forest (900–1500 m)	SF	1485	1° 54′ S, 27° 35′ E
23	Swamp Forest	SwF	2623	1° 49′ N, 17° 39′ E
24	Sparse Grassland	SG	4578	29° 18′ S, 18° 47′ E
25	Salt Hardpans	SH	1785	33° 43′ N, 8° 28′ E
26	Tree Crops	TC	154	30° 37′ N, 30° 39′ E

GLC2000 was used to select the initial land cover classes in the SPOT VGT images. Although there are 27 land cover classes for the African Continent present in GLC2000, in this work only 26 were used (Table 2) because the class 'Waterbodies' was discarded as it is a very easy class, with a spectral signature clearly different from the others. Figure 1 presents the geographical locations of the training areas, and the number of pixels for each class.

2.2. *Satellite imagery and data processing*

The VGT sensor, on board SPOT satellites, has provided daily coverage of the entire Earth since 1998, at a spatial resolution of 1 km (VITO 2012). The sensor acquires data in four spectral bands, BLUE (0.43–0.47 μm), RED (0.61–0.68 μm), NIR (0.78–0.89 μm), and SWIR (1.58–1.75 μm) (VITO 2012). Three types of SPOT VGT product are available: primary products, daily (S1), and 10 day (S10) syntheses. NDVI, which is the difference between the NIR and RED bands divided by the sum of these two bands (e.g. Tucker 1980), is computed from the S10 images.

SPOT VGT S10 syntheses are made available for ten possible regions of interest, one of these pre-defined regions being Africa. All SPOT VGT S10 composites of Africa used in this work were transferred from the SPOT VGT site (VITO 2012) and geo-referenced in

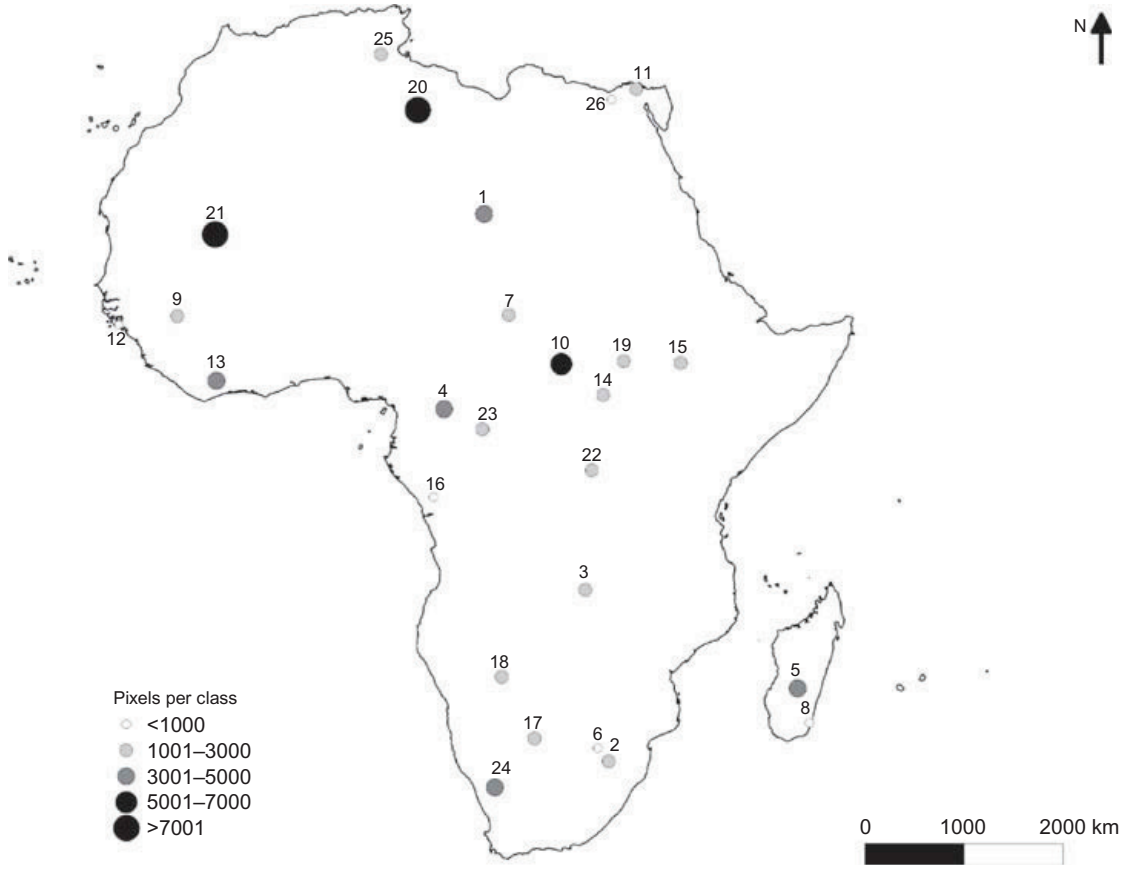


Figure 1. Geographical location of the training areas identified for each land cover class.

LAT/LONG WGS84 coordinates. For a single year (1 January–31 December), a total of 36 SPOT VGT S10 images are available (3 for each month).

As the NDVI S10 images data set from SPOT VGT images is affected by various factors such as atmospheric conditions, maximum value composites (MVC) are used to minimize their influence. Thus, for the period of 10 days, the NDVI values are analysed on a pixel-by-pixel basis and the highest value is retained for each pixel location. A MVC image is obtained when all pixels have been evaluated. Tucker, Hielkema, and Roffey (1985) showed that MVC imagery is highly related to the green vegetation dynamic. Common problems encountered in single-date remote-sensing studies, as cloud contamination, atmospheric attenuation, and observation geometry, are minimized only using MVC. However, the MVC process as applied to the NDVI S10 images data set is not sufficient to eliminate all unrealistic variability from NDVI time-series (Jonsson and Eklundh 2004; Rodrigues, Marçal, and Cunha 2013). In order to reduce the impact of additional noise introduced by the process of overlapping several images, a median filter was applied to the NDVI images used for training. The median filter smoothes the data, removing the undesired abrupt changes without introducing artefacts, or suppresses natural variation in the vegetation growth.

NDVI annual profiles were created for each class from the 36 NDVI median values extracted from each year. To remove rare events such as rapid transitions or other anomalies present in the temporal profiles, a Savitzky–Golay filter (Press et al. 2007) was applied. Savitzky–Golay filter coefficients are derived by performing an unweighted linear least square fit using a polynomial of a given degree. In general, higher-degree polynomials can capture more accurately the heights and widths of narrow peaks, but perform poorly at smoothing wider peaks (Orfanidis 1995). Some tests were done and a Savitzky–Golay

filter with a first-degree polynomial and frame size 5 showed the best results, removing the undesirable data from NDVI time-series and capturing efficiently all the transitions related with the main and double growth seasons.

The correlation between smoothed and original profiles for all 26 classes was highly significant ($r = 0.85$, $n = 36$, $p < 0.0001$). This result was interpreted as indicating that the Savitzky–Golay filter does not significantly alter the original NDVI profile, smoothing the curve and improving subsequent data analysis.

Figure 2 shows the NDVI time-series obtained for each class after application of the Savitzky–Golay filter. Four classes (BR, SDD, SD, and SH) are clearly different from all others, due their low NDVI values, but they are very similar to each other. The other classes all have particular features that can be used for the classification process.

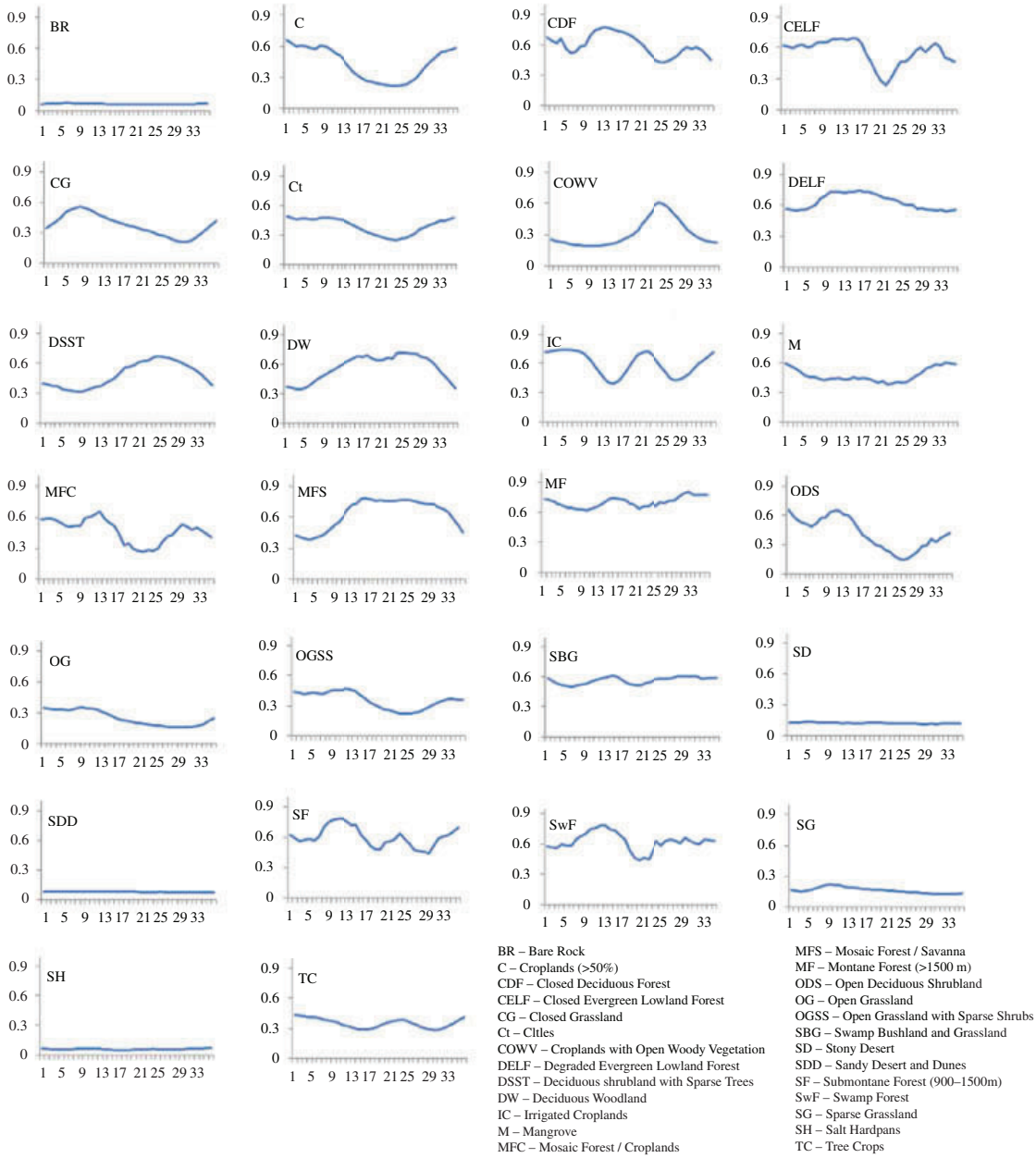


Figure 2. NDVI time-series of the 26 land cover classes for Africa, 2000, after the application of a Savitzky–Golay filter with first-degree polynomial and frame size 5.

2.3. Hierarchical aggregation

The land cover classification of a pixel makes use of its NDVI annual profile (time-series). The observed profile is compared with those of each class provided by training, using a similarity measure, and the closest is considered the class most likely. In this work, four different similarity measures were used: Euclidean distance (ED), cosine similarity measure (Cos), Jaccard coefficient (Jac), and dynamic time warping (DTW). The first three represent time-rigid measures while DTW is a time-flexible measure. To test the impacts of temporal distortions on vegetation profiles over the years, DTW was used in two modes, considering windows restrictions in the time-axis of 20 and 30 days.

2.3.1. Time-rigid similarity measures

Euclidean distance is the most widely used similarity measure to compare two time-series. Considering two time-series $\boldsymbol{T} = t_1, \ldots, t_n$ and $\boldsymbol{S} = s_1, \ldots, s_m$, where n and m represent the length of $\boldsymbol{T}$ and $\boldsymbol{S}$, respectively, the ED between $\boldsymbol{T}$ and $\boldsymbol{S}$ is defined by the square root of the sum of squared differences between corresponding elements $\boldsymbol{T}$ and $\boldsymbol{S}$ (Agrawal, Faloutsos, and Swami 1993), according to Equation (1). The lower the value of ED, the greater the similarity between $\boldsymbol{T}$ and $\boldsymbol{S}$. One of the disadvantages of this similarity measure is the fact that only time-series with the same length can be compared. In our case, this was not a problem, as only yearly time-series ($n = m = 36$) were compared:

$$ED\,(T,S) = \sqrt{\sum_{i=1}^{n} (t_i - s_i)^2},\ \textit{where } n = m. \tag{1}$$

The cosine similarity measure (Cos) defines the cosine of the angle between $\boldsymbol{T}$ and $\boldsymbol{S}$. This allows points with equal composition but different scales to be treated identically. The Cos between $\boldsymbol{T}$ and $\boldsymbol{S}$, as defined in Equation (2) (Egghe 2009), ranges from –1 (meaning exactly opposite) to 1 (meaning exactly the same). A value of zero indicates no correlation, while intermediate values indicate various levels of similarity (]0,1[) or dissimilarity (]–1,0[):

$$\mathrm{Cos}\,(T,S) = \frac{\sum_{i=1}^{n} t_i \times s_i}{\sqrt{\sum_{i=1}^{n} t_i^2} \times \sqrt{\sum_{i=1}^{n} s_i^2}}. \tag{2}$$

Cos similarity may be extended to the Jaccard coefficient (Jac). The Jaccard coefficient calculates the ratio between the intersection and union of $\boldsymbol{T}$ and $\boldsymbol{S}$ (Jaccard 1912), as shown in Equation (3) (Egghe 2009). The results range from 0 (minimum similarity) to 1 (maximum similarity):

$$\mathrm{Jac}\,(T,S) = \frac{\sum_{i=1}^{n} t_i \times s_i}{\sum_{i=1}^{n} t_i^2 + \sum_{i=1}^{n} s_i^2 - \sqrt{\sum_{i=1}^{n} t_i} \times \sqrt{\sum_{i=1}^{n} s_i}}. \tag{3}$$

2.3.2. Time-flexible similarity measures

ED, Cos, and Jac similarity measures do not consider shifts or distortions in time, which is often an important aspect when comparing temporal vegetation data. DTW allows an ‘elastic’ transformation of the time-series, minimizing the effects of delays and distortions in time, to detect similar patterns (Berndt and Clifford 1994).

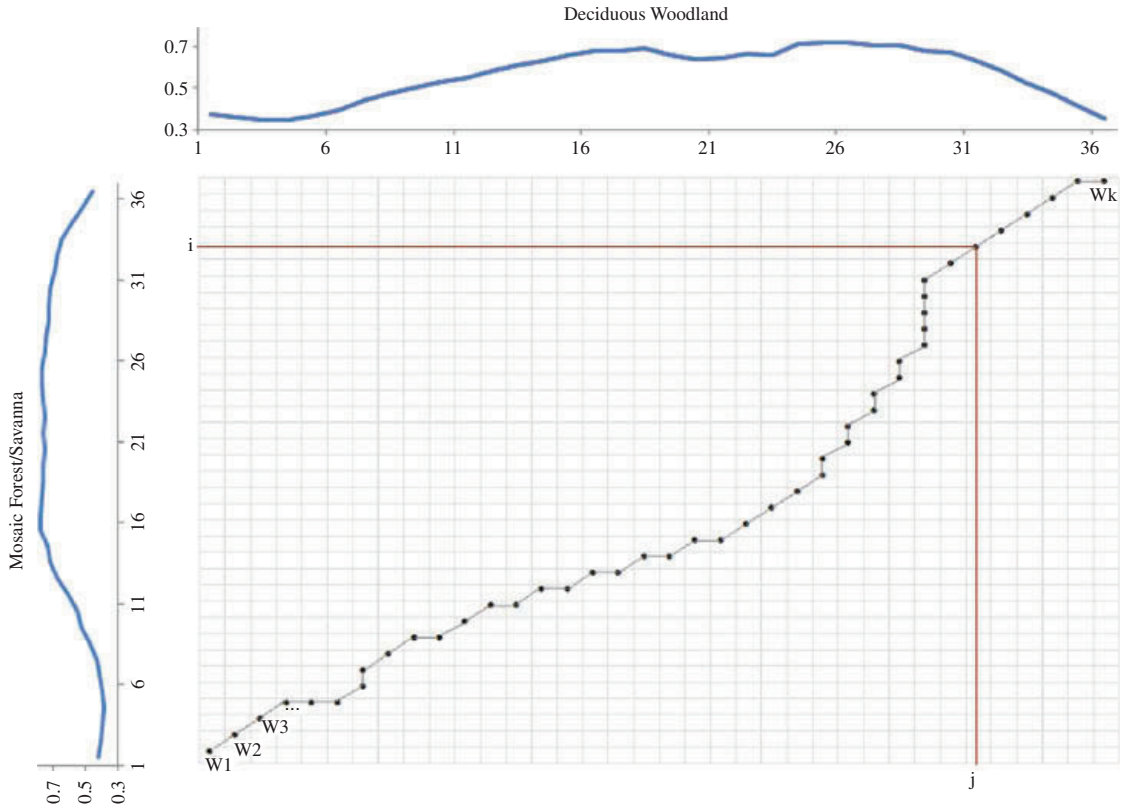


Figure 3. Example of a warping path, obtained using the classes DW (Deciduous Woodland) and MFS (Mosaic Forest/Savanna).

To determine DTW between $\boldsymbol{T}$ and $\boldsymbol{S}$, first an n-by-m matrix is created, where the $(i_{\text{th}}, j_{\text{th}})$ element represents the distance (e.g. the ED) between points t_i and s_j. This distance matrix is commonly called cost function (cf). Next, the point – to – point alignment and matching relationship between $\boldsymbol{T}$ and $\boldsymbol{S}$ can be represented by time – warping paths $\boldsymbol{W} = (w1, w2, \ldots, wk)$, being $k\,(\max n, m \leq k \leq m + n - 1)$ the number of all possible warping paths. Usually the warping path is subject to various restrictions such as boundary conditions, monotonicity, and continuity (Keogh and Pazzani 2001). A warping path is a contiguous set of matrix elements that defines a map between $\boldsymbol{T}$ and $\boldsymbol{S}$, where the element $w_k = i,j$ indicates the alignment and matching relationship between t_i and s_j (Li, Chen, and Wu 2010). As an illustration, Figure 3 presents an example of the optimal warping path obtained using the classes DW (Deciduous Woodland) and MFS (Mosaic Forest/Savanna).

The DTW between the time-series $\boldsymbol{T}$ and $\boldsymbol{S}$ is the value that minimizes Equation (4) (Chu et al. 2002). The k in the denominator is used to compensate for the fact that warping paths may have different lengths:

$$\text{DTW}\,(T,S) = min\left\{\frac{1}{k}\sqrt{\sum\nolimits_{i=1}^{k} W_i}\right\}. \tag{4}$$

When the two time-series have the same length ($n = m$), ED can be seen as a special case of DTW with no warping along the time-axis. Thus, the kth element of $\boldsymbol{W}$ is constrained such that $w_k = (i,j)_k$, with $i = j = k$ (Assent et al. 2009).

DTW can produce undesirable results if displacements and deformations are not limited in the time-axis. To avoid this, some restrictions can be applied that limit the possible time-axis difference between the two time-series (Sakoe and Chiba 1978). Various methods have been proposed, such as windowing (Berndt and Clifford 1994), slope weighting (Kruskall and Liberman 1983; Sakoe and Chiba 1978), and slope constraints (Itakura 1975; Myers, Rabiner, and Roseneberg 1980). In this work, to test the impacts of temporal distortions in the vegetation profiles over the years, two versions of DTW were used considering windows restrictions in the time-axis (shifts) of 20 (DTW2) and 30 (DTW3) days. These restrictions were chosen because, in general, interannual vegetation dynamics related to phenological stages do not present differences greater than 1 month (30 days). Figure 4 shows the effect of these restrictions in the warping path calculation, using the classes DW (Deciduous Woodland) and MFS (Mosaic Forest/Savanna) as an example. In this case, the two restrictions considered result in two different warping paths.

2.3.3. *Hierarchical aggregation methods*

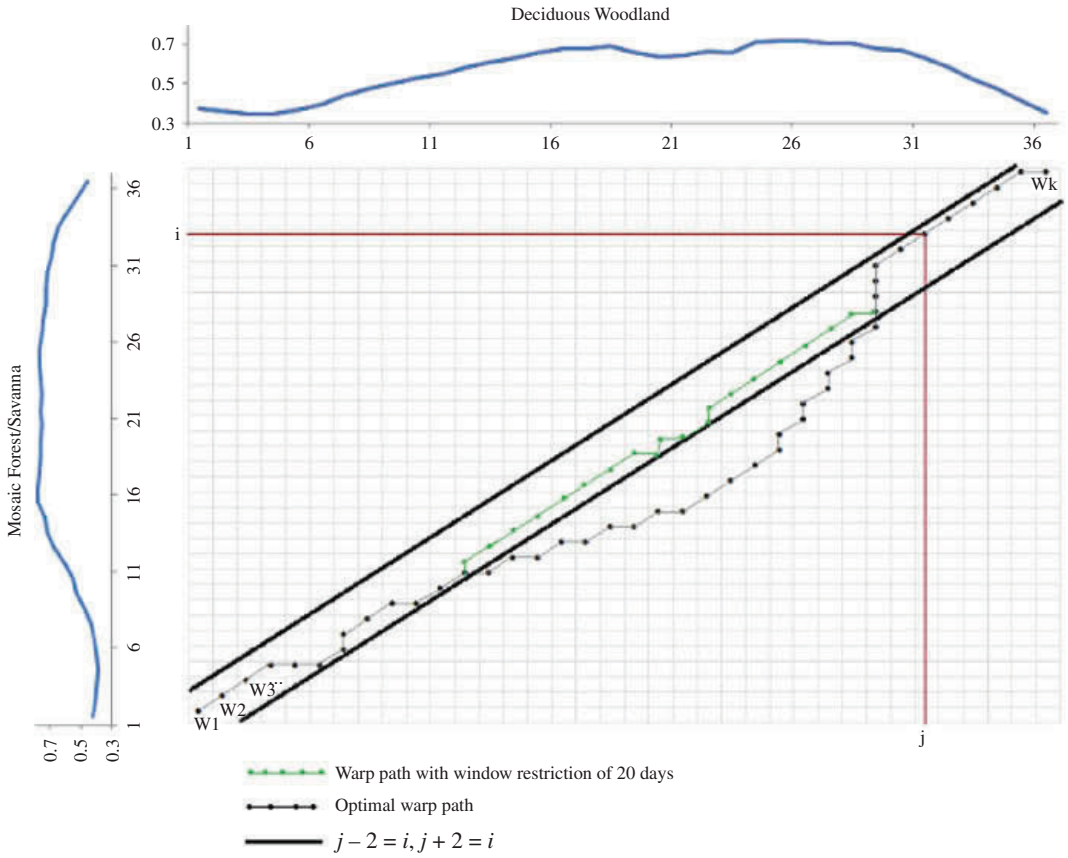
The similarity measures were applied to the filtered NDVI time-series to perform a hierarchical aggregation. This process consists in joining the two most similar classes in succession until only a single class is left. The hierarchical process has 26 levels of aggregation, from the 1st (26 classes) to the 26th (with only 1 class).

At each level of aggregation, one new class is created based on the signatures of the two classes merged from the previous level. Two distinct methods were used to characterize a new class: a New Signature (NS) and a Preserved Signature (PS). In the first case (NS), the signature of the new class is created based on the mean of the signatures (NDVI time-series) of the two previously joined, repeating the classification process at each hierarchical level with the revised set of signatures. In the second case (PS), the original signatures of the classes are preserved at all hierarchical levels and the classification process is performed only once, at the 1st level.

A cross-validation study was conducted. Cross-validation consists of partitioning the initial data into independent training and control samples, intended to produce reliable error estimates. The available reference data (Figure 1, Table 2) for each of the 26 classes were split randomly in five subsets, which were used as training (80%) and control (20%) samples. The classification process was repeated five times, leaving a different subset for control at each time.

For each class, and for each level of aggregation, the classification agreement was determined based on the number of pixels correctly classified in that class. As the PS method only performs the classification process at the 1st level, this procedure to determine the agreement was only applied at this level. For the other levels, the agreement estimation was based on the error matrix generated for the 1st level. This error matrix contains, for each class, the number of pixels correctly classified in that class and the number of pixels classified in other classes.

An example of an error matrix using four test classes is presented in Figure 5(*a*). In this example, for class 1, only 293 of a total of 960 pixels were correctly classified in that class; the others pixels were classified in classes 3 (303 pixels) and 4 (364). For class 2, all pixels were correctly classified obtaining a classification agreement of 100%. This can be explained by the fact that temporal profiles for the test classes 1, 3, and 4 are very similar (Figure 5(*b*)), in contrast to class 2 that has a clearly identifiable profile.



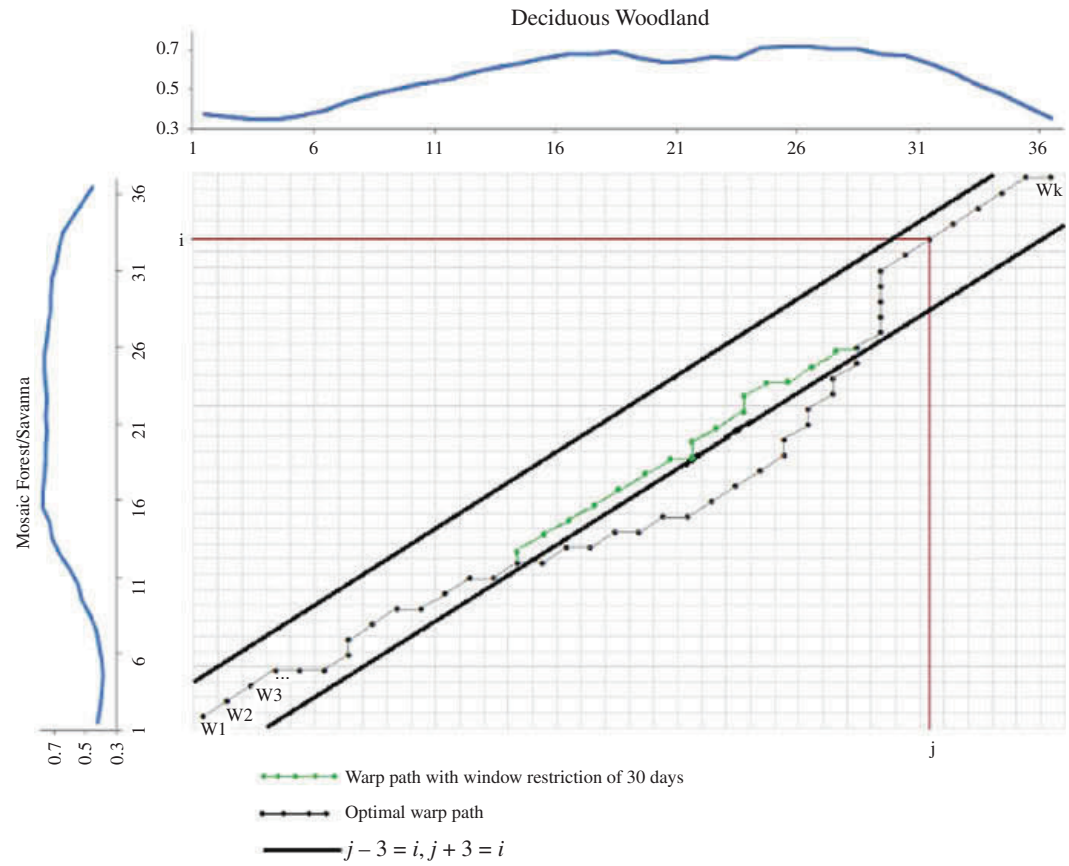
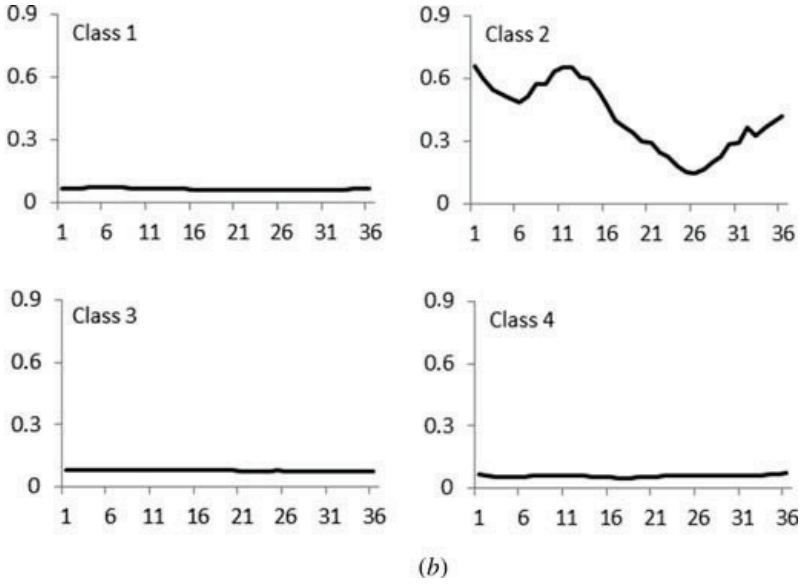


Figure 4. Example of warping paths obtained using the classes DW (Deciduous Woodland) and MFS (Mosaic Forest/Savanna) with windows restrictions of 20 (top) and 30 days (bottom).

	1	2	3	4
1	293	0	303	364
2	0	37	0	0
3	363	0	5912	0
4	76	0	7	275

(*a*)



(*b*)

Figure 5. (*a*) Example of an error matrix obtained using four distinct classes and (*b*) the temporal profiles for each class.

2.4. GLC update for multiple years

Using the year 2000 as reference, NDVI profiles for the years 1999, 2001, 2008, and 2010 were processed to evaluate the influence of temporal variability in the land cover results. These years were selected because adjacent years (1999–2000 and 2000–2001) were expected to have little change in land cover, and the differences identified between the reference and study years would also be due to greater variability in the weather, soil moisture, vegetation conditions, and interannual crop selection. In the years 2008 and 2010 the land cover changes were expected to be higher and, consequently, the classification process would have to deal with the factors previously mentioned, as well as land cover variability itself.

The same training areas (same pixel/geographical locations) used for 2000 were selected in the NDVI SPOT VGT images for 1999, 2001, 2008, and 2010. The extraction and processing of the NDVI profiles for these years were done using the same procedure as described for 2000. To identify the areas of potential change for 1999, 2001, 2008, and 2010, two different approaches were considered. In the first, the same procedure described for the year 2000 was conducted in 1999, 2001, 2008, and 2010, training year-by-year. This approach assumes that land cover in the training areas remains unchanged over the years. This assumption is obviously weak, particularly for 2008 and 2010. A second approach was also considered, which uses training samples from the year 2000 and control samples from the test year (1999, 2001, 2008, or 2010). Although this second approach allows the elimination of problems caused by differences in annual land cover in training areas, it is vulnerable to changes associated with atmospheric conditions and land cover variability over the years.

3. Results and discussion

3.1. GLC update for the reference year (2000)

Figure 6 presents the classified images for the year 2000, using Cos, ED, Jac, DTW2, and DTW3 as classifiers. Comparing the five classified images, the Cos measure shows the

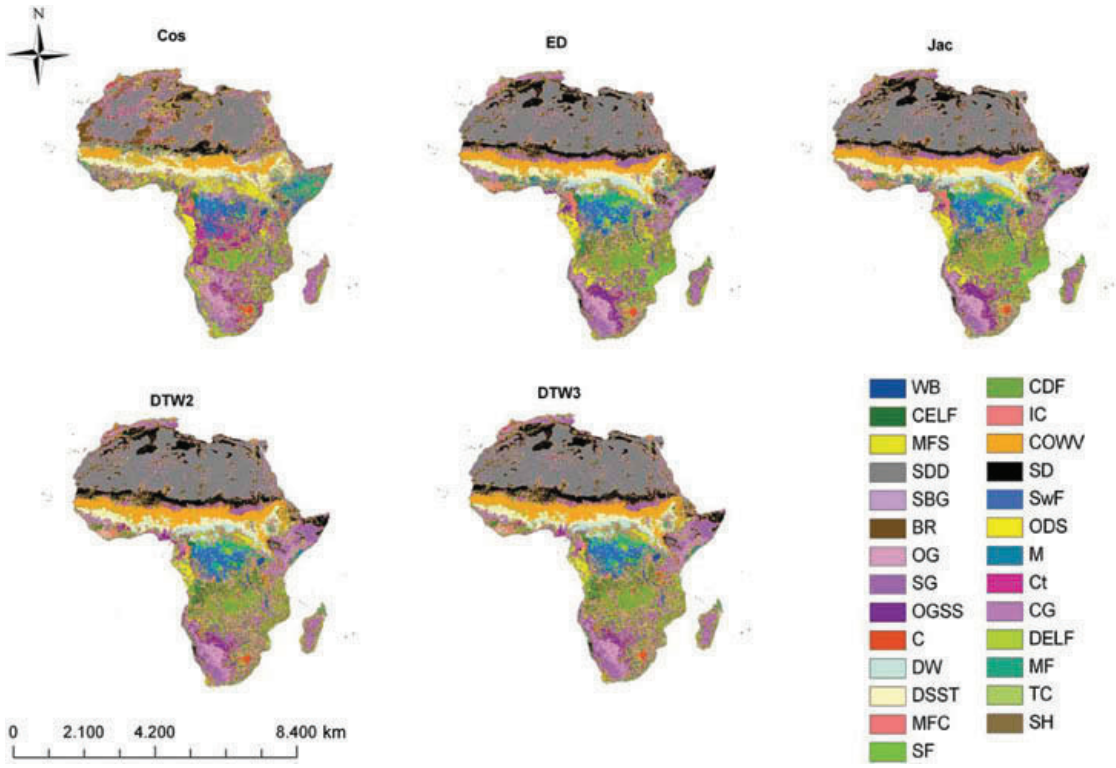


Figure 6. Classified images for the year 2000 obtained using the cosine (Cos), Euclidean Distance (ED), Jaccard (Jac), and dynamic time warping (with windows restrictions of 20 (DTW2) and 30 days (DTW3)) measures.

most distinct results, particularly in Northern Africa. Although the DTW2, DTW3, ED, and Jac images present similar results for the majority of the continent, there are clearly distinguishable differences in Western Africa.

The average agreement (Table 3) between the five similarity measures only slightly varies, from 89.8% to 93.3%, but there are considerable differences for individual classes. The best performance is obtained using Cos (93.3% on average), with a minimum of 73.6% (for Salt Hardpans (SH)). The results obtained using DTW2 and DTW3 were no better than when using other similarity measures, which can be explained by the fact that training and control data from the same year are being used.

The classes Cities (Ct), Mangrove (M), and Bare Rock (BR) had lower classification agreement, except for the Cos measure. In contrast, the classes Croplands with Open Woody Vegetation (COWV), Open Deciduous Shrubland (ODS), and Stony Desert (SD) gave a classification agreement of 100% (or nearly) for all similarity measures.

A comparison between the average classification agreements obtained for each level of aggregation using NS and PS methods is presented in Figure 7, with best similarity measure (Cos). In the first ten hierarchical levels the two methods gave very similar results, with NS slightly outperforming PS. For higher aggregation levels (12–24) the results of the two methods show wide disparity, with the PS method performing best. The variability in the NS results can be justified by modification of spectral signatures at each level.

The average agreement versus aggregation-level plots showed similar behaviour for ED, Jac, DTW2, and DTW3 to those presented in Figure 7 for Cos.

Table 3. Classification agreement (%) for each class and similarity measures for 2000.

		Classification agreement (%)				
Class	Code	ED	Jac	Cos	DTW2	DTW3
1	BR	31.6	31.8	77.4	53.2	88.9
2	C	99.7	99.6	99.9	99.9	99.8
3	CDF	99.9	99.8	99.5	98.3	97.1
4	CELF	98.1	97.6	95.4	99.1	99.2
5	CG	98.7	99.3	99.8	98.6	99.4
6	Ct	34.7	33.9	95.7	29.9	29.2
7	COWV	100.0	100.0	99.9	100.0	100.0
8	DELF	93.3	96.3	100.0	92.6	90.7
9	DSST	99.3	99.5	99.8	96.7	96.2
10	DW	99.8	99.8	95.7	99.2	99.5
11	IC	98.5	98.8	96.0	97.8	97.6
12	M	56.8	56.8	99.3	47.5	46.0
13	MFC	99.6	99.6	86.9	98.9	98.2
14	MFS	99.9	100.0	99.2	99.2	98.9
15	MF	93.0	95.6	83.4	93.6	92.8
16	ODS	100.0	100.0	100.0	100.0	100.0
17	OG	99.4	99.7	96.5	99.7	99.8
18	OGSS	92.8	91.9	81.5	82.4	81.6
19	SBG	92.1	90.7	96.4	87.0	86.4
20	SD	100.0	100.0	99.4	100.0	100.0
21	SDD	94.3	94.5	89.5	97.8	97.9
22	SF	98.0	97.8	96.4	97.8	97.2
23	SwF	98.6	98.7	99.4	96.3	95.2
24	SG	98.3	98.6	87.4	98.8	99.0
25	SH	77.8	77.4	73.6	93.4	93.9
26	TC	91.6	90.9	77.4	83.1	83.8
Average agreement (%)		90.2	90.3	93.3	90.0	89.8

Notes: The classification agreement for each class represents the number of pixels correctly classified in that class, and the average agreement is calculated using the average of the 26 values. ED: Euclidean distance; Jac: Jaccard Coefficient; Cos: cosine measure; DTW2: dynamic time warping with a restriction in time-axis of 20 days; DTW3: dynamic time warping with a restriction in time-axis of 30 days.

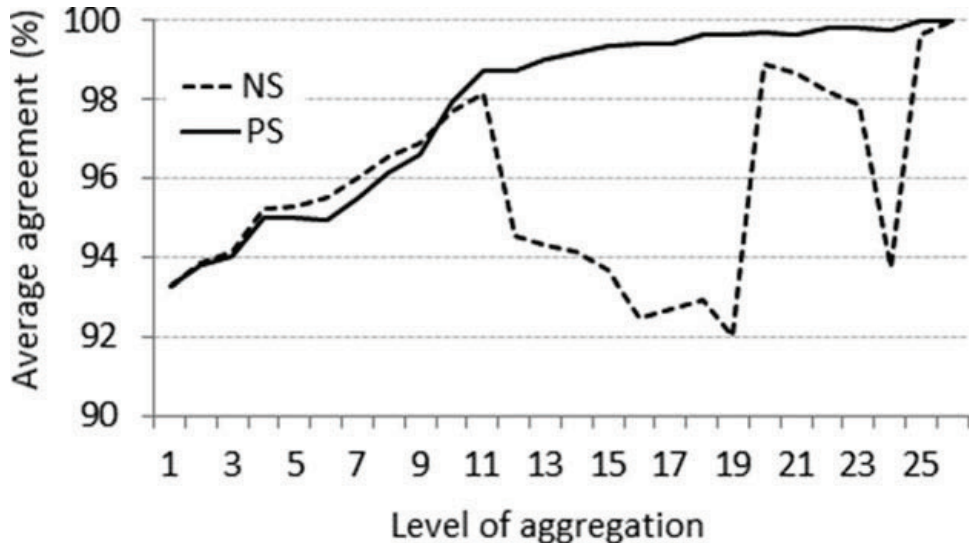


Figure 7. Average classification agreement (%) as a function of aggregation level, obtained from NS and PS methods, at each level of aggregation, using the cosine measure.

3.2. GLC update for multiple years

The study done for the year 2000 was repeated for 1999, 2001, 2008, and 2010. To evaluate the proposed methodology for these years, two approaches were considered: (1) using the training and control samples from the same year and (2) using the training samples and signatures from 2000 and the control samples from the year under evaluation (1999, 2001, 2008, or 2010).

3.2.1. *Training and control samples from the same test year*

The training and control samples selected for the reference year (2000) were used with 1999, 2001, 2008, and 2010 data. This is based on the assumption that no land cover changes had occurred at the test sites between that year and 2000. Table 4 presents the results obtained for the different years, using the similarity measures for levels 1 (26 classes), 7 (20 classes), and 12 (15 classes) using the two distinct aggregation methods (PS and NS).

For the first level, the values obtained for the years 1999, 2001, 2008, and 2010 are not very different to those obtained for 2000, with the Cos measure being the best performer.

Table 4. Average classification agreement (%) using training and validation data from the year under evaluation.

		Average agreement (%)				
		26 classes	20 classes		15 classes	
Year	Similarity measure	NS = PS	NS	PS	NS	PS
1999	ED	90.3	93.7	93.6	94.4	95.0
	Jac	90.4	93.5	93.7	96.3	96.5
	Cos	91.6	96.2	95.8	97.9	98.4
	DTW2	89.2	90.5	90.7	91.6	93.2
	DTW3	89.3	90.3	90.5	91.4	93.0
2000	ED	90.2	96.1	95.9	95.5	95.5
	Jac	90.3	96.2	96.0	98.0	98.4
	Cos	93.3	96.0	95.5	94.5	98.7
	DTW2	90.0	93.8	94.1	92.2	93.3
	DTW3	89.8	93.2	93.8	92.5	93.1
2001	ED	91.1	96.1	96.7	94.9	97.2
	Jac	91.2	94.1	94.8	96.1	97.9
	Cos	94.5	94.7	96.6	93.7	97.7
	DTW2	90.9	93.5	93.3	94.5	96.1
	DTW3	90.0	92.2	93.2	90.6	91.9
2008	ED	88.7	92.6	93.1	93.9	93.8
	Jac	89.0	92.5	92.6	96.1	97.1
	Cos	91.1	91.9	93.0	91.8	95.3
	DTW2	87.1	90.1	90.6	93.3	96.8
	DTW3	86.8	88.9	89.7	90.7	91.9
2010	ED	86.7	92.3	92.4	92.3	93.6
	Jac	86.9	90.5	91.2	95.2	96.8
	Cos	89.1	92.1	91.0	92.5	92.6
	DTW2	88.2	90.6	90.8	92.7	95.2
	DTW3	88.0	90.1	90.4	91.9	95.0

The lowest average agreement (87%) was obtained for 2008 and 2010, which are the years most distant to the reference year (2000).

A comparison between NS and PS methods was done for the hierarchical levels with 15 and 20 classes. The results show, in general, an increase in the average agreement through the aggregation levels, for both PS and NS. With 20 classes the difference between the two methods is no greater than 1.9%, but with 15 classes the difference is larger, and in some cases greater than 4% (e.g. in 2000 and 2001 using the Cos measure), with PS performing best.

The results presented in Table 4 show that a reduction in the number of classes from 26 to 15 provides classification agreements above 90.5%, for all similarity measures, with a better performance for the PS method.

The results of the hierarchical aggregation process, obtained from the Cos measure for the years 2000, 2001, and 2010, are presented in Figures 8, 9, and 10. The dendogram analysis for the reference year (Figure 8) shows that it is possible to identify seven groups of classes. The classes assigned to each group and a description of the group characteristics are presented in Table 5. For the years 2001 and 2010 (Figures 9 and 10) it is possible to see some differences in the aggregation of classes. For 2001, the classes with more differences were those of the fourth group (Table 5). This group includes the grass, shrubs, and agricultural vegetation types. Weather conditions, water resources, crop management, grass cutting, and grazing are among the factors that can affect the growth period of these land cover types, which makes them more susceptible to changes over time. For the year 2010, classes 8 (Degraded Evergreen Lowland Forest), 11 (Irrigated Croplands), and 26 (Tree Crops) showed differences in the aggregation process when compared with the year 2000. These classes represent water-dependent and degraded vegetation types, and are thus more likely to show different NDVI profiles over the years.

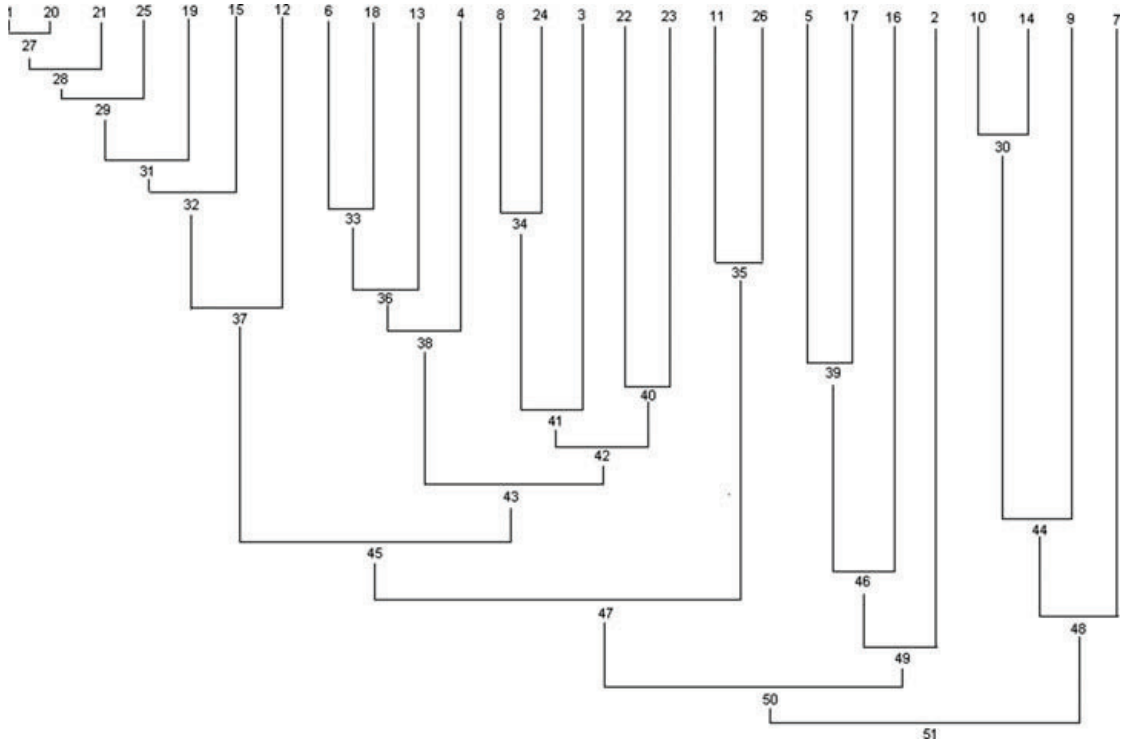


Figure 8. Dendogram for the hierarchical aggregation process, obtained from the cosine measure for 2000.

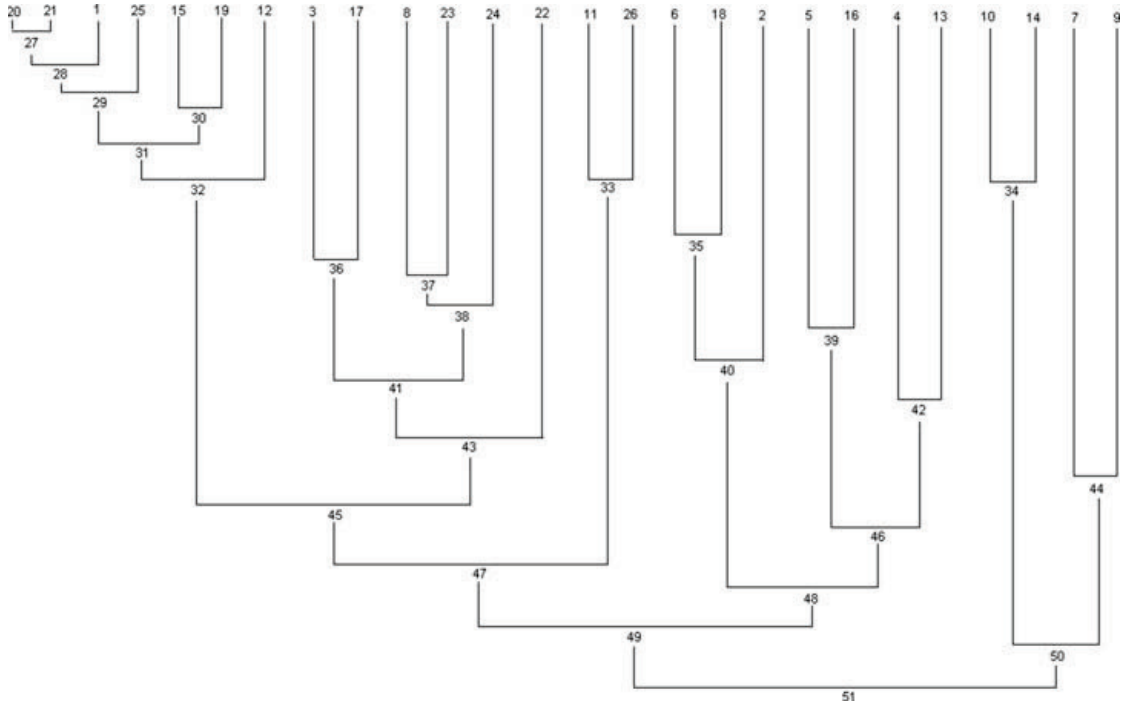


Figure 9. Dendogram for the hierarchical aggregation process, obtained from the cosine measure for 2001.

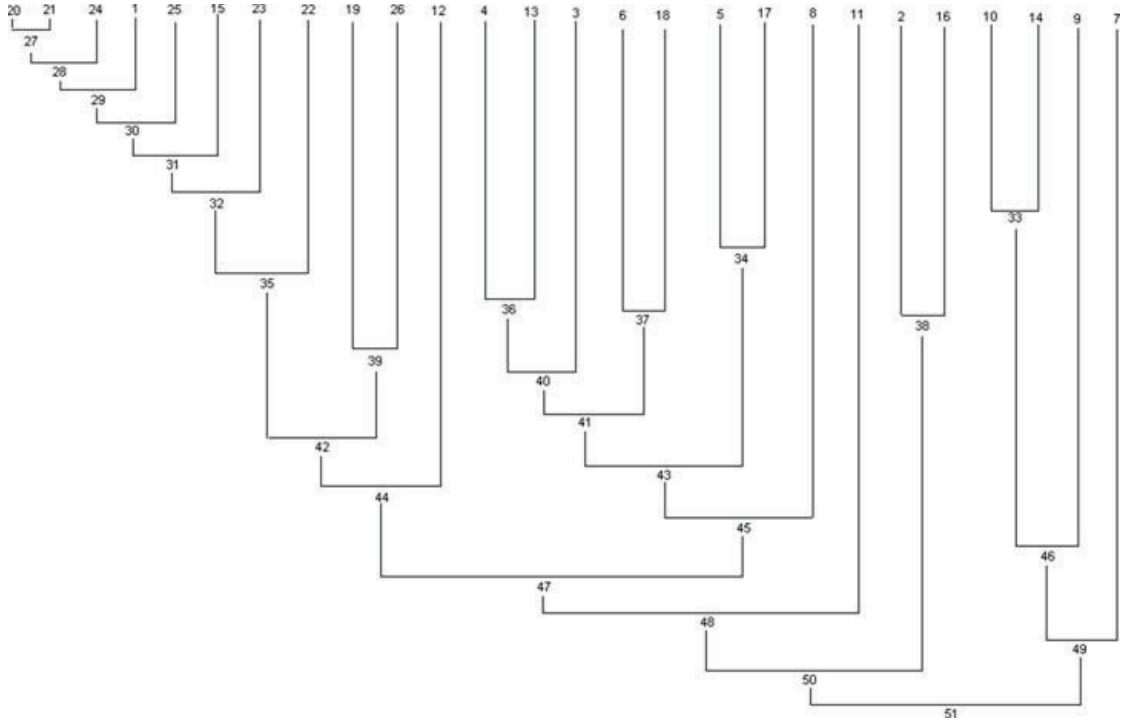


Figure 10. Dendogram for the hierarchical aggregation process, obtained from the cosine measure for 2010.

The classes showing annual regrowth (e.g. classes 2 (Croplands > 50%), 11 (Irrigated Croplands), and 16 (Open Deciduous Shrubland)) can be more susceptible to changes over the years, as they are influenced by different factors such as weather conditions, pasture management, and available water resources.

Table 5. Description of the seven groups of classes identified through the dendogram analysis, for the year 2000.

Group	Classes	Description
1	BR, SD, SDD, SH, MF, SBG, M	Presents an almost linear NDVI profile over the year, which is a characteristic of non-vegetated or low-vegetation areas.
2	Ct, OGSS, MFC, CELF, DELF, SG, CDF, SF, SwF	Maximum vegetation development around April/May.
3	IC, TC	Maximum vegetation development in July, verifying an autumn regrowth period.
4	CG, OG, ODS, C	NDVI maximum development between March and April, followed by a decrease in the NDVI until autumn.
5	DW, MFS	Growth period from February to May, stabilizing in summer.
6	DSST	Regular profile during summer, achieving peak growth in September.
7	COWV	NDVI peak in August, followed by brief senescence and dormancy.

3.2.2. Training in 2000 and control samples from other years

In spite of the results presented in the previous section, the fixed-year approach is somehow limited because it is based on the assumption that the land cover used for training from GLC2000 is unchanged for 1999, 2001, 2008, and 2010. To resolve this problem, an alternative approach was considered, using training data from 2000 (reference year) and control samples from other years (1999, 2001, 2008, or 2010).

As expected, the classification agreement values obtained with this approach (Table 6) are worse than those obtained in the previous section. This can be explained by the fact that training and control samples from different years support different dynamics of vegetation influenced by weather and other temporal variability. Figure 11 presents an example of an error matrix obtained from the application of the ED measure, for the cross-validation process, using training data from 2000 and control samples (subset 1) from 1999. Note that classes 16 (Open Deciduous Shrubland) and 17 (Open Grassland) have a small number of pixels correctly classified, being those most classified in classes 5 (Closed Grassland) and 24 (Sparse Grassland). This is very different to the results obtained for the reference year, where classes 16 and 17 achieved classification agreements of 100% and 99.2%, respectively. In order to better understand these differences, Figure 12 presents the NDVI profile of classes 5 (CG), 16 (ODS), 17 (OG), and 24 (SG) obtained for 1999 and 2000. These profiles show the changes that occurred between 1999 and 2000, particularly for the Open Deciduous Shrubland and Closed Grassland classes. The Sparse Grassland and Open Grassland classes present a similar annual profile, in particular for the year 1999, where they are almost coincident over days 180–330. This similarity can explain the results showed in the error matrix (Figure 11) and consequently the reduction in classification agreement. The same problem was observed with other classes for the years 2001, 2008, and 2010, which proves that temporal variability influences vegetation profiles over the years.

Due to variation in interannual vegetation dynamics, it was expected that DTW2 and DTW3, which use a flexible time-axis perspective and can thus adapt to intra-annual climate

Table 6. Average classification agreement (%) obtained from all years, using training data from 2000.

Year	Similarity measure	Average agreement (%)				
		26 classes	20 classes		15 classes	
		NS = PS	NS	PS	NS	PS
1999	ED	64.9	73.1	71.3	78.2	69.3
	Jac	65.2	68.7	67.9	74.7	74.5
	Cos	54.6	57.8	57.1	62.8	62.4
	DTW2	70.4	74.6	76.0	78.8	79.1
	DTW3	69.8	75.2	75.9	79.3	78.7
2000	ED	90.2	96.1	95.9	95.5	95.5
	Jac	90.3	96.2	96.0	98.0	98.4
	Cos	93.3	96.0	95.5	94.5	98.7
	DTW2	90.0	93.8	94.1	92.2	93.3
	DTW3	89.8	93.2	93.8	92.5	93.1
2001	ED	72.9	80.5	80.3	88.1	84.5
	Jac	74.1	82.8	82.1	90.3	92.1
	Cos	52.2	58.8	57.8	65.9	67.0
	DTW2	70.5	78.8	79.4	83.0	85.0
	DTW3	71.0	78.7	79.9	83.3	83.9
2008	ED	62.5	70.9	70.0	73.4	72.5
	Jac	62.0	69.1	68.6	80.6	79.8
	Cos	51.4	55.0	52.5	57.2	54.7
	DTW2	67.2	75.7	76.0	76.7	79.0
	DTW3	67.5	74.3	75.3	76.6	78.7
2010	ED	59.6	69.3	69.6	74.1	72.9
	Jac	58.3	65.5	65.4	76.4	75.6
	Cos	41.0	50.5	49.2	53.7	50.6
	DTW2	68.7	74.9	76.7	81.0	81.4
	DTW3	68.2	74.1	75.8	80.2	80.7

variability, would produce better results than the fixed-time similarity measures. The results confirm this hypothesis, except for 2001. In general, the greater the gap between the training and control years the greater the temporal variability, and thus the results tend to be worse (Table 6). In general, for both PS and NS methods, average agreement increases with aggregation level, showing a similar pattern to that verified in the previous section. When the number of classes was reduced from 26 to 15, average agreement was improved, and this was verified when using only annual data (Table 4).

As we are comparing and using data from different years, it is important to evaluate the time-rigid (ED, Jac, and Cos) and the time-flexible similarity measures (DTW2 and DTW3). As DTW and ED are the most commonly used time-series similarity measures (Agrawal et al. 1995; Das et al. 1998; Debregeas and Hebrail 1998; Faloutsos, Ranganathan, and Manolopoulos 1994; Keogh et al. 2001), a comparison between these two measures was carried out. Both DTW2 and DTW3 were tested, but as their results were found to be very similar, only DTW2 was considered in comparison with ED. Table 7 presents a comparison between the average classification agreement (%) obtained from DTW2 and ED, for 26 classes, using annual and interannual data. For both similarity measures, it will be seen

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	242	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	328	0	0	0	390	0
2	0	301	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0
3	0	0	124	0	0	0	0	54	0	0	0	0	0	0	0	0	0	0	0	0	0	284	20	0	0	0
4	0	0	8	245	0	0	0	0	0	0	0	8	0	0	0	0	0	0	33	0	0	8	404	0	0	0
5	0	0	0	0	748	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0
6	0	2	0	0	0	38	0	0	0	0	0	0	11	0	0	0	30	33	0	0	0	0	0	2	0	0
7	0	0	0	0	0	0	430	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	1	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	0	3	0	345	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1121	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	1	0	0	0	0	0	554	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1
12	0	0	0	4	0	4	0	0	0	0	0	15	0	0	0	0	0	0	3	0	0	0	1	0	0	1
13	0	2	53	291	0	0	0	0	0	0	0	221	83	0	0	0	0	0	112	0	0	1	2	0	0	0
14	0	0	0	0	0	0	0	0	0	1	0	0	0	533	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	198	0	0	0	18	0	0	0	8	0	0	0
16	0	0	0	0	34	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	10	0	0	0	577	0	0
18	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	340	250	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	1	0	0	0	12	0	3	24	0	0	0	275	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1588	0	0	0	0	0	0
21	365	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5910	0	0	0	0	0
22	0	0	3	0	0	0	0	85	0	0	0	0	0	0	0	0	0	0	0	0	0	171	38	0	0	0
23	0	0	0	4	0	0	0	65	0	0	0	0	0	0	0	0	0	0	0	0	0	54	402	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	715	0	0	0	201	0	0
25	108	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	237	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	28

Figure 11. Error matrix for the Euclidean distance (ED) measure using training of 2000 and control samples of 1999, for a subset of the cross-validation process.

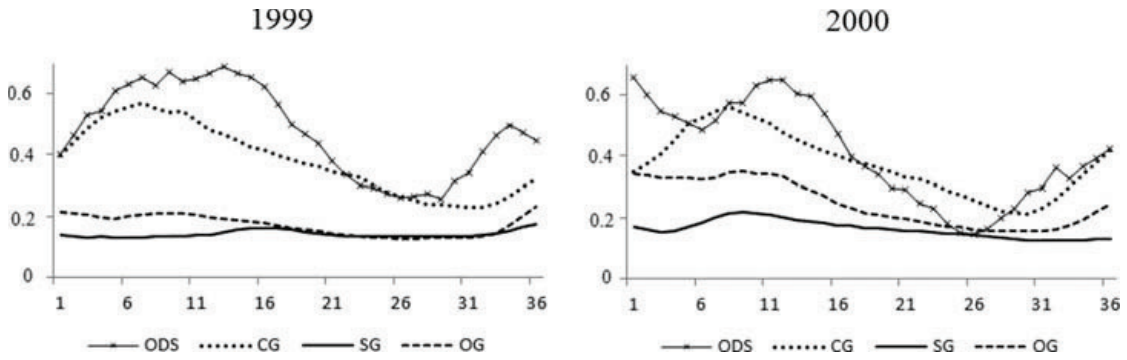


Figure 12. NDVI temporal profiles for Closed Grassland (CG), Open Deciduous Shrubland (ODS), Open Grassland (OG), and Sparse Grassland (SG), for the years 1999 and 2000.

that, in general, the more distant the year being analysed from the reference year (2000), the worse the results. The differences between DTW2 and ED are reinforced in the interannual approach. The use of interannual data can introduce errors in the classification process, because the vegetation dynamic for a specific class can change over the years influenced by uncontrollable conditions such as anthropogenic actions, agricultural practices, available soil water, and weather. DTW2 can detect these time differences if they occur in a small bandwidth (20 days), which is not possible using ED. Although the differences between the time-flexible and time-rigid similarity measures in an annual perspective are low ($<1.6\%$), analysis of the interannual approach showed marked differences, varying from 2.4% in 2001 to 9.1% in 2010.

Table 7. Average classification agreement (%) obtained from DTW2 and ED, for 26 classes, using annual and interannual data.

Training data	Year	ED	DTW2	Δ
In the year under evaluation	1999	90.3	89.2	1.1
	2001	91.1	90.9	0.2
	2008	88.7	87.1	1.6
	2010	86.7	88.2	1.5
Reference year	2000	90.2	90.0	0.2
In 2000	1999	64.9	70.4	5.5
	2001	72.9	70.5	2.4
	2008	62.5	67.2	4.7
	2010	59.6	68.7	9.1

Note: Δ is the absolute difference between ED and DTW2 values.

4. Conclusions

This article presents a methodology to identify areas of potential change on a continental scale, based on NDVI time-series extracted from SPOT VGT. The method was illustrated as applied to Africa based on the GLC2000 data set and using 26 classes. A hierarchical aggregation structure for the land cover classes was computed based on the similarity of the NDVI profiles. For each level of aggregation, the average classification agreement for each class was calculated. This methodology proved to be very useful in identifying the areas of potential change at various levels of class aggregation, and thus determining the best selection for the number of classes, maintaining the desirable level of agreement.

The classification agreements were obtained in relation to the GLC2000 data set, which has its own inaccuracies. Independent ground truth data are therefore needed to confirm that this approach is robust and accurate, and thus able to be used to update land cover maps for different years, facilitating the study of interannual land cover changes. The methodology proposed is not restricted to SPOT VGT data, and can be used for any temporal data set.

Acknowledgements

The authors would like to thank VITO for the SPOT VEGETATION images, and also the Global Vegetation Monitoring Unit of the European Commission Joint Research Centre for providing access to GLC2000.

This study was carried out under a Doctoral Grant (SFRH/BD/62189/2009) funded by the Fundação para a Ciência e a Tecnologia (FCT).

References

Agrawal, R., C. Faloutsos, and A. Swami. 1993. "Efficient Similarity Search in Sequence Databases." *Lectures Notes in Computer Science* 730: 69–84.

Agrawal, R., K. I. Lin, H. S. Sawhney, and K. Shim. 1995. "Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Times-Series Databases." In *Proceedings of the 21st International Conference of Very Large Data Bases*, edited by U. Dayal, P. M. D. Gray, and S. Nishio, 490–501, Zurich, September 11–15. Burlington, MA: Morgan Kaufmann. ISBN 1-55860-379-4.

Assent, I., M. Wichterich, R. Krieger, H. Kremer, and T. Seidl. 2009. "Anticipatory DTW for Efficient Similarity Search in Time Series Databases." *Proceedings of Very Large Data Base Endowment* 2: 826–837.

Bagan, H., P. J. Wang, and Y. Yasuoka. 2007. "Cropland Area Extraction in China with Multi-Temporal MODIS Data." *American Geophysical Union*, Fall Meeting 2007.

Bartholomé, E., and A. S. Belward. 2005. "GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data." *International Journal of Remote Sensing* 26: 1959–1977.

Berndt, D. J., and J. Clifford. 1994. "Using Dynamic Time Warping to find patterns in time series." In *Knowledge Discovery in Databases: Papers from the 1994 AAAI Workshop*, edited by U. M. Fayyad and R. Uthurusamy, 359–370, Seattle, WA, July 1994. Technical Report WS-94-03. Palo Alto, CA: AAAI Press. ISBN 0-929280-73-3.

Bosard, M., J. Feranec, and J. Otahel. 2000. *CORINE Land Cover Technical Guide—Addendum 2000*. European Environmental Agency Technical Report No. 40. Copenhagen.

Carrão, H., P. Gonçalves, and M. Caetano. 2007. "Use of Intra-Annual Satellite Imagery Time-Series for Land Cover Characterization Purposes." *EARSeL eProceedings* 6: 1–11.

Chu, S., E. Keogh, D. Hart, and M. Pazzani. 2002. "Iterative Deepening Dynamic Time Warping for Time Series." In *Proceedings of the Second SIAM International Conference on Data Mining*, edited by R. L. Grossman, J. Han, V. Kumar, H. Mannila, R. Motwani, 195–212, Arlington, VA, April 11–13. SIAM. ISBN 0-89871-517-2.

Clark, M. L., T. M. Aide, H. R. Grau, and G. Riner. 2010. "A Scalable Approach to Mapping Annual Land Cover at 250 m Using MODIS Time-Series Data: A Case Study in the Dry Chaco Ecoregion of South America." *Remote Sensing of Environment* 114: 2816–2832.

Colditz, R. 2007. "Time Series Generation and Classification of MODIS Data for Land Cover Mapping." Dissertation Document, Institute of Geography and Geology, University of Wurzburg.

Das, G., K. Lin, H. Mannila, G. Renganathan, and P. Smyth. 1998. "Rule Discovery from Time Series." In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*, edited by R. Agrawal, P. E. Stolorz, G. Piatetsky-Shapiro, 16–22, New York, August 27–31. Palo Alto, CA: AAAI Press. ISBN 1-57735-070-7.

Debregeas, A., and G. Hebrail. 1998. "Interactive Interpretation of Kohonen Maps Applied to Curves." In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*, edited by R. Agrawal, P. E. Stolorz, G. Piatetsky-Shapiro, 179–183, New York, August 27–31. Palo Alto, CA: AAAI Press. ISBN 1-57735-070-7.

DeFries, R. S., M. C. Hansen, J. R. G. Townshend, and R. S. Sohlberg. 1998. "Global Land Cover Classifications at 8 Km Spatial Resolution: The Use of Training Data Derived from Landsat Imagery in Decision Tree Classifiers." *International Journal of Remote Sensing* 19: 3141–3168.

DeFries, R. S., and J. R. G. Townshend. 1994. "NDVI-Derived Land Cover Classifications at a Global Scale." *International Journal of Remote Sensing* 15: 3567–3586.

Egghe, L. 2009. "New Relations between Similarity Measures for Vectors Based on Vector Norms." *Journal of the American Society for Information Science and Technology* 60: 232–239.

Faloutsos, C., M. Ranganathan, and Y. Manolopoulos. 1994. "Fast Subsequence Matching in Time-Series Databases." *Association for Computing Machinery Special Interest Group on Management of Data Record* 23: 419–429.

Friedl, M. A., D. K. McIver, J. C. F. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf. 2002. "Global Land Cover Mapping from MODIS: Algorithms and Early Results." *Remote Sensing of Environment* 83: 287–302.

Geisser, S. 1975. "The Predictive Sample Reuse Method with Applications." *Journal of the American Statistical Association* 70: 320–328.

Gumma, M. K., P. S. Thenkabail, and A. Nelson. 2011. "Mapping Irrigated Areas Using MODIS 250 Meter Time-Series Data: A Study on Krishna River Basin (India)." *Water* 3: 113–131.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, and R. A. Sohlberg. 2000. "Global Land Cover Classification at 1 km Spatial Resolution Using a Classification Tree Approach." *International Journal of Remote Sensing* 21: 1331–1364.

Heremans, S., B. Bossyns, H. Eerens, and J. V. Orshoven. 2011. "Efficient Collection of Training Data for Sub-Pixel Land Cover Classification Using Neural Networks." *International Journal of Applied Earth Observation and Geoinformation* 13: 657–667.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. "Development of a 2001 National Land-Cover Database for the United States." *Photogrammetric Engineering and Remote Sensing* 70: 829–840.

Hopfner, C., and D. Scherer. 2011. "Analysis of Vegetation and Land Cover Dynamics in North-Western Morocco during the Last Decade Using MODIS NDVI Time Series Data." *Biogeosciences Discussions* 8: 3953–3998.

Itakura, F. 1975. "Minimum Prediction Residual Principle Applied to Speech Recognition." *IEEE Transactions on Acoustics Speech and Signal Processing* ASSP-23: 52–72.

Jaccard, P. 1912. "The Distribution of the Flora in the Alpine Zone." *The New Phytologist* IX: 37–50.

Jonsson, P., and L. Eklundh. 2004. "TIMESAT - A Program for Analyzing Time-Series of Satellite Sensor Data." *Computers & Geosciences* 30: 833–845.

Keogh, E., K. Chakrabarti, M. Pazzani, and S. Mehrotra. 2001. "Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases." *Knowledge and Information Systems* 3: 263–286.

Keogh, E., and M. Pazzani. 2001. "Derivative Dynamic Time Warping." Proceedings of the First SIAM International Conference on Data Mining, Chicago, IL, April 5–7.

Kleynhans, W., J. C. Olivier, K. J. Wessels, B. P. Salmon, F. van den Bergh, and K. Steenkamp. 2011. "Detecting Land Cover Change Using an Extended Kalman Filter on MODIS NDVI Time-Series Data." *IEEE Geoscience and Remote Sensing Letters* 8: 506–511.

Kruskall, J. B., and M. Liberman. 1983. *Time Warps, String Edits and Macromolecules: The Theory and Practice of String Comparison: The Symmetric Time Warping Algorithm: From Continuous to Discrete*. Reading, MA: Addison-Wesley.

Landmann, T., M. Schramm, R. R. Colditz, A. Dietz, and S. Dech. 2010. "Wide Area Wetland Mapping in Semi-Arid Africa Using 250-Meter MODIS Metrics and Topographic Variables." *Remote Sensing* 2: 1751–1766.

Latifovic, R., and D. Pouliot. 2005. "Multitemporal Land Cover Mapping for Canada: Methodology and Products." *Canadian Journal of Remote Sensing* 31: 347–363.

Li, Y., H. Chen, and Z. Wu. 2010. "Dynamic Time Warping Distance Method for Similarity Test of Multipoint Ground Field." *Mathematical Problems in Engineering* 2010: 1–12. Article ID 749517.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant. 2000. "Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1 Km AVHRR Data." *International Journal of Remote Sensing* 21: 1303–1330.

Lunetta, R. S., J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy. 2006. "Land-Cover Change Detection Using Multi-Temporal MODIS NDVI Data." *Remote Sensing of Environment* 105: 142–154.

Machwitz, M., T. Landman, A. Cord, C. Conrad, and S. Dech. 2008. "Land Cover Analysis on Sub-Continental Scale: FAOLCSS Standard with 250 Meter MODIS Satellite Observations in West Africa." Geoscience and Remote Sensing Symposium, V49–V52, Boston, MA, July 7–11. ISBN 978-1-4244-2808-3.

Marçal, A. R. S., J. S. Borges, J. A. Gomes, and J. F. Pinto da Costa. 2005. "Land Cover Update by Supervised Classification of Segmented ASTER Images." *International Journal of Remote Sensing* 26: 1347–1362.

Mayaux, P., E. Bartholomé, S. Fritz, and A. Belward. 2004. "A New Land-Cover Map of Africa for the Year 2000." *Journal of Biogeography* 31: 861–877.

Myers, C., L. Rabiner, and A. Roseneberg. 1980. "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition." *IEEE Transactions on Acoustics Speech and Signal Processing* ASSP-28: 623–635.

Orfanidis, S. J. 1995. *Introduction to Signal Processing*. Uppper saddle River, NJ: Prentice-Hall.

Pôças, I., M. Cunha, A. R. S. Marçal, and L. S. Pereira. 2011. "An Evaluation of Changes in a Mountainous Rural Landscape of Northeast Portugal Using Remotely Sensed Data." *Landscape and Urban Planning* 101: 253–261.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 2007. *Numerical Recipes: The Art of Scientific Computing*, 766–772. Cambridge: Cambridge University Press.

Rodrigues, A., A. R. S. Marçal, and M. Cunha. 2013. "Monitoring Vegetation Dynamics Inferred by Satellite Data Using the PhenoSat Tool." *IEEE Transactions on Geoscience and Remote Sensing* 51: 2096–2104.

Sakoe, H., and S. Chiba. 1978. "Dynamic Programming Algorithm Optimization for Spoken Word Recognition." *IEEE Transactions on Acoustics Speech and Signal Processing* 26: 43–49.

Stone, M. 1974. "Cross-Validatory Choice and the Assessment of Statistical Predictions (with Discussion)." *Journal of the Royal Statistical Society B* 36: 111–147.

Tucker, C. J. 1980. "Remote Sensing of Leaf Water Content in the Near Infrared." *Remote Sensing of Environment* 10: 23–32.

Tucker, C. J., J. U. Hielkema, and J. Roffey. 1985. "The Potential of Satellite Remote Sensing of Ecological Conditions for Survey and Forecasting Desert-Locust Activity." *International Journal of Remote Sensing* 6: 127–138.

Vanacker, V., M. Linderman, F. Lupo, S. Flasse, and E. Lambin. 2005. "Impact of Short-Term Rainfall Fluctuation on Interannual Land Cover Change in Sub-Saharan Africa." *Global Ecology and Biogeography* 14: 123–135.

Vintrou, E., A. Desbrosse, A. Bégué, S. Traoré, C. Baron, and D. Lo Seen. 2012. "Crop Area Mapping in West Africa Using Landscape Stratification of MODIS Time Series and Comparison with Existing Global Land Products." *International Journal of Applied Earth Observation and Geoinformation* 14: 83–93.

VITO (Free SPOT VEGETATION Products). 2012. Accessed February 30, 2012. http://free.vgt.vito.be/

Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and J. N. Van Driel. 2001. "Completion of the 1990's National Land Cover Data Set for the Conterminous United States." *Photogrammetric Engineering and Remote Sensing* 67: 650–662.

Wardlow, B. D., and S. L. Egbert. 2008. "Large-Area Crop Mapping Using Time-Series MODIS 250 m NDVI Data: An Assessment for the U.S. Central Great Plains." *Remote Sensing of Environment* 112: 1096–1116.

Zhan, X., R. A. Sohlberg, J. R. G. Townshend, C. DiMiceli, M. L. Carroll, J. C. Eastman, M. C. Hansen, and R. S. DeFries. 2002. "Detection of Land Cover Changes Using MODIS 250 m Data." *Remote Sensing of Environment* 83: 336–350.

LAND COVER MAP PRODUCTION FOR BRAZILIAN AMAZON USING NDVI SPOT VEGETATION TIME SERIES

1. Summary

The Brazilian Amazon forest is a study area for many classification research studies because of its unique characteristics and interfaces with other important vegetation types. The landscape transformations and deforestation since the 1970s have been contributing to the increase of atmospheric carbon dioxide concentration, with severe present and future climate consequences (Batjes and Sombroek, 1997; Houghton, 2000; Schulze et al., 2003). Thus up-to-date and accurate land cover maps are required for understanding and monitoring the impacts of these changes in the environment (soil degradation, biotic diversity, carbon cycle). However, the production of up-to-date land cover maps in the tropical regions is difficult due to the presence of cloud cover and/or due to the difficulty in acquiring ground data.

Aiming to solve these limitations, the methodology proposed in the previous chapter was applied in the Brazilian Amazon. It is important to note the difficulty in classifying a large number of vegetated classes over the Amazon, due the cloud contamination (particularly during the rainy season) and the characteristics of the region. Two major problems that usually affect the classification accuracy were evaluated: the effect of the similarity between the land cover classes and the land cover variability over the years.

Can. J. Remote Sensing, Vol. 39, No. 4, pp. 277–289, 2013

Land cover map production for Brazilian Amazon using NDVI SPOT VEGETATION time series

A. Rodrigues, A.R.S. Marcal, D. Furlan, M.V. Ballester, and M. Cunha

Abstract. Earth Observation Satellite (EOS) data have a great potential for land cover mapping, which is mostly based on high resolution images. However, in tropical areas the use of these images is seriously limited due to the presence of clouds. This paper evaluates the ability of temporal-based image classification methods to produce land cover maps in tropical regions. A new approach is proposed for land cover classification and updating based exclusively on temporal series data, illustrated with a practical test using SPOT VEGETATION satellite images from 1999 to 2011 for Rondonia (Amazon), Brazil. Using the GLC2000 as reference, a Normalized Difference Vegetation Index (NDVI) time series of 15 distinct land cover classes (LCC) were created. Two classifiers were used (Euclidean Distance and Dynamic Time Warping) to produce maps of land cover changes for 1999–2011. Due to the difficulties in discriminating 15 LCC in the Amazon region, a hierarchical aggregation was performed by joining the initial classes gradually up to four broad classes. The land cover changes in the 1999–2011 period were evaluated using criteria based on the classification results for the individual years. The comparison with reference data showed consistent results, proving that this approach is able to produce accurate land cover maps using exclusively temporal series EOS data.

Résumé. Les données des Satellites d'Observation de la Terre ont un grand potentiel pour la cartographie du couvert végétal, la plupart basée sur des images d' haute résolution. Cependant, l'utilisation de ces images en régions tropicales est sérieusement limitée en raison de la présence de nuages. Ce document évalue l'adéquation des méthodes de classification en utilisant des images temporelles pour produire des cartes d'occupation des sols dans les régions tropicales. Une nouvelle approche est proposée pour classification de la couverture terrestre et mises à jour basée uniquement sur des données de séries temporelles, et illustrée par un test pratique en utilisant des données du satellite SPOT VEGETATION entre 1999 et 2011 pour Rondonia (Amazonie), Brésil. En prenant comme référence le GLC2000, ont été créés séries temporelles de NDVI pour 15 différents types de couverture terrestre (TCT). Deux classificateurs ont été utilisé (Distance Euclidienne et « Dynamic Time Warping ») pour produire cartes des modifications de la couverture du sol pour le période 1999–2011. Due de la difficulté en classifier 15 TCT en la région de l'Amazonie, une agrégation hiérarchique a été faite en joignant les classes initiales graduellement jusqu'à quatre vaste classes. Les modifications de la couverture terrestre au période de 1999–2011 ont été évaluées par un critère basé sur les résultats de la classification pour chaque année. La comparaison avec les données de référence a montré résultats conformes, ce qui prouve que cette approche est capable de produire exact cartes d'occupation du sol en utilisant exclusivement des données de séries temporelles.

Introduction

The global environmental change research community requires improved and up to date land cover maps at regional to global scales to support a variety of science and policy applications. This is particularly true in the Brazilian Amazon where land cover change induced either by human and natural causes has been unprecedented in recent decades (Alves et al., 2009; Asner et al., 2009; Ballester et al., 2003; Braswell et al., 2003; Brown et al., 2007; Houghton et al., 2000; Li et al., 2012).

While the need for frequent monitoring of land cover changes is clear, it is difficult to produce these maps using only ground-based information (Loveland et al., 2000). The automatic classification of Earth Observation Satellite (EOS) data is thus frequently used for land cover mapping, mostly based on high resolution image data (e.g., from Landsat satellites). The ability to apply automatic classification methods to produce land cover change maps and information, with the range, quality, and detail needed by scientists and resource management, is a huge challenge. This is emphasized in tropical areas where the presence of cloud cover seriously limits the use of passive EOS data (Carreiras et al., 2003). In the case of the Amazon, cloud cover is frequent particularly in the rainy season. To eliminate this limitation, the use of low spatial resolution

Received 30 September 2012. Accepted 31 May 2013. Published on the Web at http://pubs.casi.ca/journal/cjrs on 30 October 2013.

A. Rodrigues[1] and M. Cunha. Universidade do Porto, Faculdade de Ciências, Centro de Investigação em Ciências Geo-Espaciais – CICGE, Porto, Portugal.

A.R.S. Marcal. Universidade do Porto, Faculdade de Ciências, Departamento de Matemática and INESC TEC, Porto, Portugal.

D. Furlan and M.V. Ballester. Universidade de São Paulo – USP, Centro de Energia Nuclear na Agricultura – CENA, Piracicaba, Brasil.

[1]Corresponding author: (e-mail: dr.arlete@gmail.com).

EOS data that have a very high acquisition frequency (daily) such as AVHRR (Loveland et al., 2000), SPOT VEGETATION (Carreiras et al., 2003), and MODIS (Wardlow and Egbert, 2008) can be an alternative. However, to better use these data sources, the classification needs to be slightly different than the standard approach. Instead of using a single image, where the multispectral values are used as features for each pixel, the temporal yearly Normalized Difference Vegetation Index (NDVI) profile is used as a feature vector instead.

The aim of this work was to propose a new approach for land cover classification and updates based exclusively on temporal series of NDVI, illustrated with a practical case using SPOT VEGETATION (VGT) satellite data from 1999 to 2011 for Rondonia (Amazon), Brazil. The Global Land Cover Map of the Year 2000 (GLC2000) of South America was used to select 15 distinct land cover classes (LCC) in the SPOT VGT images. Temporal yearly NDVI profiles were extracted for each LCC and a pixel-by-pixel classification was made using the Euclidean Distance (ED) and Dynamic Time Warping (DTW) as classifiers. Two approaches were considered: one that uses the NDVI time series of the reference year (2000) to classify all years; and another that considers the NDVI time series obtained from the year under analysis.

Because of the difficulty in classifying 15 vegetated classes over the Amazon, the 15 initial LCC were aggregated in four main groups based on their NDVI profile similarity. The ED and DTW measures were used to compute a hierarchical aggregation of the initial classes using the method proposed in Rodrigues et al. (2013). The effect of the NDVI time series similarity on the classification results was evaluated by performing a new classification using the four new groups of classes and comparing the results with the land cover maps obtained for the 15 initial classes.

The LCC variability over the years was analyzed by classifying each pixel as "permanent", "stable", "change", "stable in the first years", "stable in the last years", and "unspecified" using a criteria based on the number of identical classes obtained from 1999 to 2011. To evaluate the ED and DTW land cover maps between 2000 and 2011 a comparison was made with the PRODES data (PRODES, 2013) available for 2000 and 2011.

Materials and methods

Study area

A test area was established in northwest Brazil, containing the entire state of Rondonia. The transformation of the tropical forests to pastures and crops in western Amazonia, particularly at the Rondonia state, was rapidly developed under a series of government projects for resettling small farmers from southern Brazil (Ballester et al., 2003). These projects resulted in high deforestation rates, which have been well documented since the early 1980s until today through the use of satellite imagery (Alves et al., 1999; INPE, 2000). In this region, deforestation evolution leads to a typical spatial fishbone pattern associated with the colonization process and road development (Dale et al., 1994; Moran, 1993; Pedlowski et al., 1997; Ballester et al., 2003). Large areas of previously untouched rainforest have undergone rapid change since the 1970s due to extensive immigration and colonization. In just 40 years (1970–2010), the population grew from 116 thousand to 1.5 million inhabitants, and the state lost over 34% of its native forest cover (INPE, 2000). Spatially, the deforestation process is more intense along the highway BR364, between Cuiba and Porto Velho, that was constructed in 1965, improved in 1969 (Fujisaka et al., 1996), and finally paved in 1984. The main land use in the region is cattle pasture, rising from 67% of the state cultivated area in 1985 to 91% in 2006, when perennial and annual crops only accounted for 5% of agricultural areas of Rondonia, respectively (IBGE, 2013).

The GLC2000 dataset (Eva et al., 2004) was used to establish training sites, characterizing the most relevant land cover classes in this region. Although the GLC2000 map has problems picking up classes over vegetated land surfaces (Eva et al., 2006; Gonsamo and Chen, 2011), it was the only complete land cover map available. The GLC2000 classes were compared with ground reference information acquired for the year 2000. The field data were obtained by recording each point location using a Garmin 76-Map GPS. Each control point was photographed, and land cover at the eight cardinal points was recorded. **Figure 1** presents the GLC2000 land map for the central region of Rondonia (left) and the location of the 133 field data points (right). **Table 1** presents a comparison between the field and GLC2000 data. The main differences occurred in the urban and water classes, with good agreement for the other classes evaluated. For example, in the GLC2000 the rip and (or) regrowth field classes were assigned to the Agriculture Intensive, Mosaic Agriculture, and (or) Degraded Forest classes, which are crops also characterized by a regrowth around the months of July and August. The results in **Table 1** show a coherent correspondence between the two datasets for the majority of the pixels (approximately 88%).

The GLC2000 was thus considered as the reference land cover map. For each of the 15 LCC used, a training site was selected carefully to avoid pixel boundaries with other classes. The size of the training sites was, on average, 37 pixels, ranging from 15 to 252 pixels. **Table 2** presents the description, number of pixels, and coefficient of variation (Reed et al., 2002) for the NDVI profiles of the pixels assigned to each class. A total of 10 LCC present a coefficient of variation below 10%, with an overall mean of 11.0% for the 15 classes.

Most sites were identified in the central region of Rondonia, where in the last three decades there has been considerable change in land use and (or) land cover, mostly due to deforestation for crop and grassland production (Eva

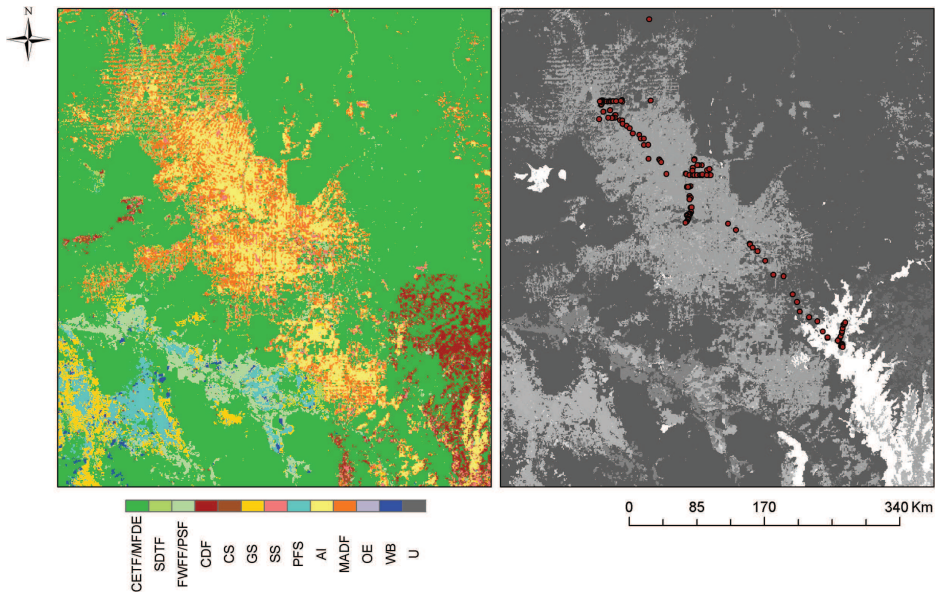


Figure 1. Global Land Cover Map (GLC) 2000 for the central region of Rondonia (left) and the location of the 133 field data points (right). The description of the GLC classes is available in **Table 2**.

et al., 2002). There are also some areas of transition between forest and regeneration, which were deforested and later abandoned, thus starting a natural regeneration process. The classes were chosen to include native vegetation, regeneration areas, intensive agriculture, flooded vegetation, and urban areas.

The Brazilian Amazon, in particular the state of Rondonia, has only two well-defined seasons: dry and rainy. The start of the dry season may vary from year to year, but it always includes the months of July and August. There is no significant variation of the air temperature throughout the year. Thus the main climatic influence for vegetation development is the annual hydric regime. The average monthly precipitation in the state capital, Porto Velho, for the year 2000 is presented in **Figure 2**. This plot clearly shows the presence of a dry season, with one month (June) almost without rain. This pluviometric pattern has a great influence in the vegetation development of grassland and shrubland and consequently on the annual pattern of NDVI values.

Satellite imagery and data preprocessing

The VGT sensor onboard the SPOT4 satellite has provided daily coverage of the entire Earth since 1998, at a spatial resolution of 1 km (VITO, 2012). It acquires data, which are delivered as primary products or daily (S1) and ten-days (S10) syntheses, in four spectral bands (blue, red, near-infrared and short-wave infrared), ranging from 0.43 to 1.75 μm (VITO, 2012). The NDVI is computed from the extracted pixel values as

$$\mathrm{NDVI} = (\rho_{\mathrm{NIR}} - \rho_{\mathrm{RED}})/(\rho_{\mathrm{NIR}} + \rho_{\mathrm{RED}}) \quad (1)$$

where ρ_{NIR} is the reflectance at the near-infrared wavelength band and ρ_{RED} is the reflectance at the red waveband (Rouse et al., 1973). The widths of the reference bands of the VGT sensor are 0.61–0.68 μm (RED) and 0.78–0.89 μm (NIR).

Ten-day NDVI synthesis images (NDVI S10) are available from VGT globally and are divided into 10 regions, one of them being South America. These synthesis images are Maximum Value Composites (MVCs) of daily NDVI images, intended to reduce the noise caused by a variety of biophysical factors. For each year, a total of 36 VGT NDVI S10 images are available (three for each month).

The entire training area for each of the 15 LCC was considered as a unit instead of using a pixel-by-pixel approach. The median of the NDVI S10 values of the pixels assigned for each class were computed. The yearly NDVI time series were created using the median values obtained for each of the 36 images available in one year.

Image classification procedure

The land cover classification of a pixel makes use of its NDVI annual profile (time series). The pixel-by-pixel classification process applied in this work compares the annual NDVI pixel profile with each of the 15 reference time series (15 NDVI class profiles). The class label of the most

Table 1. Land cover comparison between field data and Global Land Cover Map for the year 2000.

Field data		GLC2000	
Class	Pixels (no.)	Class	Pixels (no.)
Crop	26	Agriculture/mosaic agriculture	16
		Mosaic forest/dense evergreen	3
		Shrublands	7
Forest	10	Mosaic agriculture/ degraded forest	5
		Closed evergreen tropical forest	3
		Agriculture intensive	1
		Semideciduous transition forest	1
Next to forest	10	Next to forest	10
Pasture	28	Agriculture intensive	20
		Mosaic agriculture/ degraded forest	8
Rip/regrowth	33	Agriculture intensive	22
		Mosaic agriculture/ degraded forest	11
Shrublands	8	Shrub savannah	1
		Shrublands	5
		Mosaic agriculture/dense forest	2
Forest and shrub (burnt)	2	Agriculture intensive	2
Water	4	Shrub savannah	1
		Closed evergreen tropical forest	1
		Semideciduous transition forest	1
		Agriculture intensive	1
Urban	12	Urban	5
		Agriculture intensive	2
		Grass savannah	1
		Mosaic agriculture/ degraded forest	3
		Shrub savannah	1

similar reference profile is retained for the pixel under analysis. To compare the NDVI time series, the ED (Agrawal et al., 1993) and DTW (Berndt and Clifford, 1994; Chu et al., 2002) similarity measures were used. These measures were selected as they are the commonly used algorithms for similarity analysis of time series (Agrawal et al., 1995; Das et al., 1998; Debregeas and Hebrail, 1998; Faloutsos et al., 1994; Keogh et al., 2001). Furthermore, these distance measures have important differences when comparing time series data. The ED is a time-rigid measure and does not consider any distortions or shifts over time, whereas the DTW is a time-flexible measure capable to consider the shifts and distortions over time, which can have an important role in vegetation dynamics. To test the

Table 2. Characterization of the land cover classes.

Land cover description	Acronym	Pixels (no.)	Coefficient of variation (%)
Agriculture Intensive	AI	35	6.0
Closed Deciduous Forest	CDF	16	9.8
Closed Evergreen Tropical Forest	CETF	252	8.6
Closed Shrublands	CS	16	9.1
Fresh Water Flooded Forest	FWFF	48	20.3
Grass Savannah	GS	30	10.2
Mosaic Agriculture – Degraded Forest	MADF	16	7.6
Montane Forest (500–1000m) – Dense Evergreen	MFDE	24	8.5
Ore Exploration	OE	15	10.7
Periodically Flooded Savannah	PFS	16	7.4
Permanent Swamp Forest	PSF	16	5.6
SemiDeciduous Transition Forest	SDTF	16	5.8
Shrub Savannah	SS	16	7.7
Urban	U	15	12.7
Water Bodies	WB	20	35.3

impacts of temporal distortions, the DTW was applied

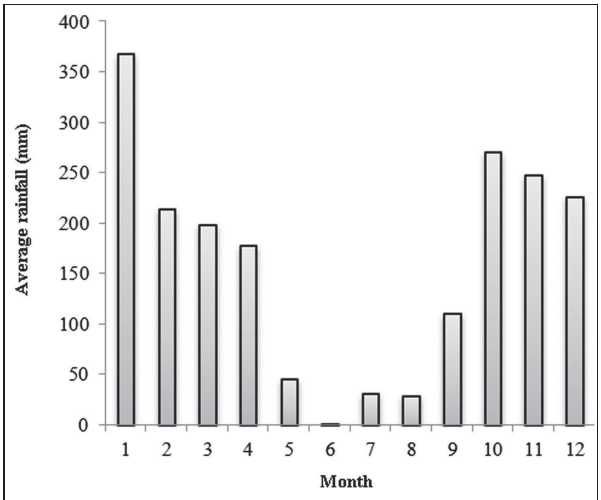


Figure 2. Average monthly rainfall (mm) for Porto Velho (Santo Antônio), for the year 2000. Source: Agência Nacional de Água (ANA).

using a window restriction of 20 days (Berndt and Clifford, 1994) in the time axis (shift).

The 15 reference LCC selected in 2000 were also selected for the remaining years (1999 and 2001–2011). For each year, two classified images were computed using the ED and DTW as classifiers. To evaluate the influence of temporal land cover variability, two approaches were considered: one that establishes the reference NDVI time series of each year

Table 3. Criteria to define an attribute for each pixel, according to the classification results from 1999 to 2011.

Classification attribute	Criteria
1 – Permanent	11+ years (out of 13) with the same class
2 – Stable	4+ in the first 5 years and 4+ in the last 5 years classified in the same class
3 – Change	4+ in the first 5 years classified in the same class Cl1, and 4+ in the last 5 years classified in the same class Cl2. Cl1 $\neq$ Cl2
4 – Stable in the first years	4+ in the first 5 years classified in the same class
5 – Stable in the last years	4+ in the last 5 years classified in the same class
6 – Unspecified	None of the previous assignments

to process the classification, and another that uses the NDVI profiles identified on the reference year (2000) to classify every year.

The effect of the NDVI time series similarity on the classification results was evaluated, reducing the number of initial LCC and analyzing the land cover variability for each pixel over the years 1999–2011. A hierarchical aggregation of the initial classes was performed using the ED and DTW distance measures, following the method proposed in Rodrigues et al. (2013), and a new classification was made using the new groups of classes. To analyze the LCC variability over years, each pixel was classified as permanent, stable, change, stable in the first years, stable in the last years, and unspecified using the criteria presented in **Table 3**. These criteria are based on the number of identical classes obtained from 1999 to 2011. **Figure 3** presents a schematic illustration of the criteria used, where each color represents a distinct LCC.

To evaluate the consistency of the ED and DTW land cover maps for 2000 and 2011 a comparison was made with the PRODES data available for 2000 and 2011. PRODES (Amazon Deforestation Monitoring Project) consists of a systematic monitoring of deforestation in Amazon, using remote sensing images and digital image processing techniques (PRODES, 2013). ED and DTW maps were reclassified in three main classes: forest, nonforest, and hydrography. The percentage of pixels classified in each class was determined and the consistency of the results was evaluated using the PRODES data as reference.

Results and discussion

NDVI S10 images from SPOT VGT from 1999 to 2011 were processed for Rondonia. As precipitation has considerable influence on the vegetation development, the most relevant period for this region is from April to September. Furthermore, the presence of clouds limits the use of satellite data in the rainy season. Maximum Value Composite images of January and April–September using NDVI VGT data from 2000 are presented in **Figure 4**. Due to the frequent cloud cover, the NDVI values from January might not be clearly related to the vegetation type and condition. In the remaining images, it was noticeable that the central areas of Rondonia have lower NDVI values. This is because this part of the state has low height herbaceous or semi-herbaceous vegetation, as the soil is mostly used for agriculture and pasture. Moreover, Rondonia is a remarkable example of land cover change in the past decades as a result of deforestation induced by human and natural causes. The converted forest cover in a fishbone pattern, mainly due to forestation caused by agricultural and urban expansion as reported by Eva et al. (2002), is noticeable in **Figure 4**. The least influence of clouds occurred between June and August. During this period a large increase in NDVI occurred, especially in forests. The better transparency of the atmosphere at this time of the year promotes a net balance of radiation greater than the other parts of the year (Da Rocha et al., 2004; Malhi et al., 2002). This high availability of energy and the ability of forestry trees to capture water from deep soil explain the trend of the NDVI increase in forest areas between June and August.

The original (dotted line) and Savitzky–Golay (Press et al., 2007) smoothed (solid line) NDVI time series training profiles of the 15 LCC, for the year 2000, are presented in **Figure 5**. Each LCC multitemporal NDVI signature is consistent with its expected pattern of vegetation dynamics

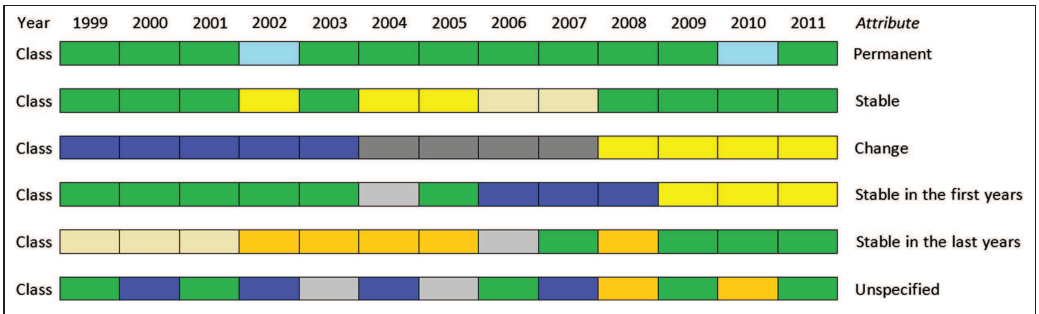


Figure 3. Schematic illustration explaining the criteria to define an attribute for each pixel, according to the classification results from 1999 to 2011. Each color represents a distinct class.

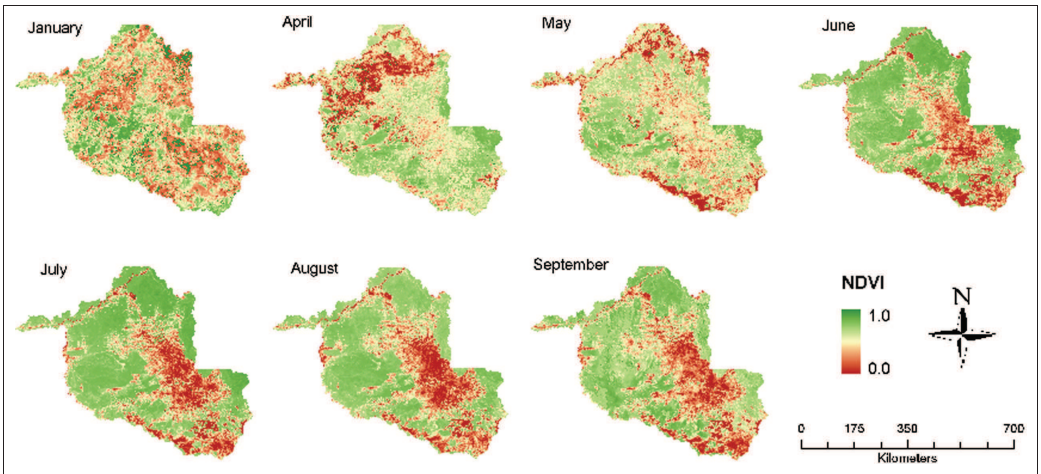


Figure 4. NDVI 10-day composites for Rondonia produced from SPOT VEGETATION data for the year 2000. The NDVI values correspond to the first 10 days of each month.

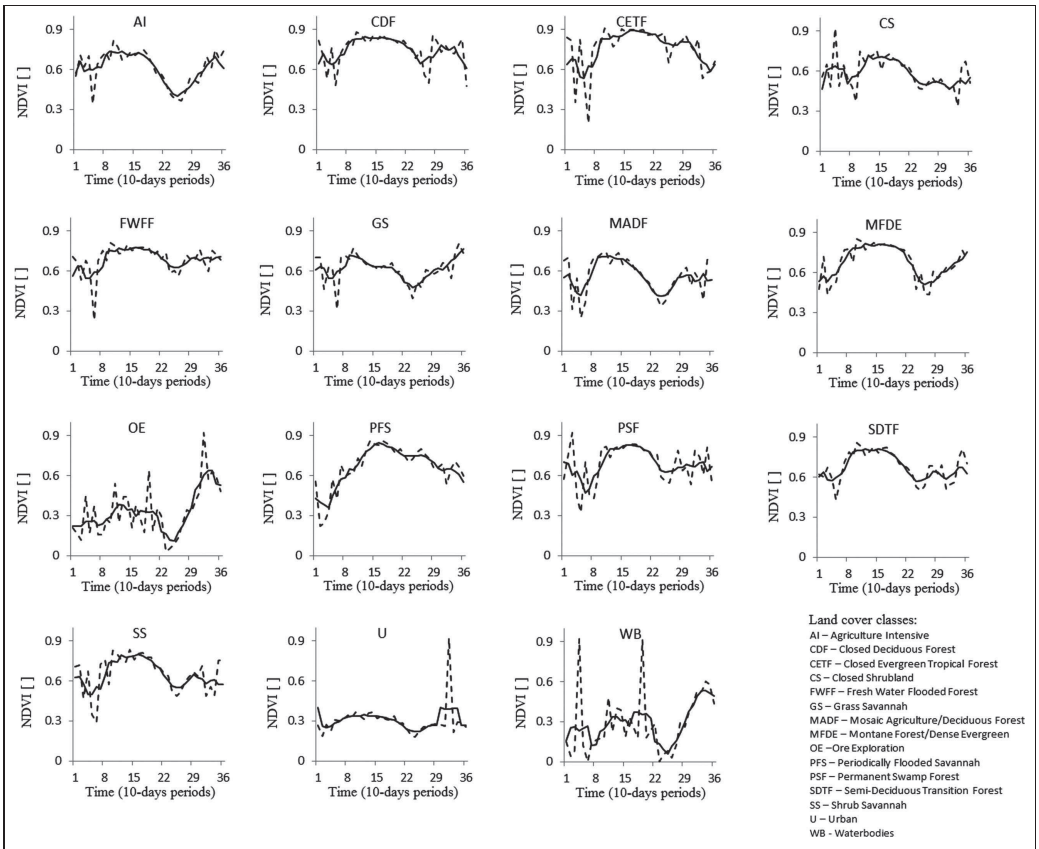


Figure 5. Original (dotted line) and Savitzky–Golay smoothed (solid line) NDVI time series of the 15 land cover classes in Rondonia for the year 2000.

and most of them are spectrally separable at some point during the growing season. The NDVI profiles consistently showed a major influence of the presence of clouds, especially from November to February. Land cover classes representing shorter vegetation types were the most sensitive to the hydrological regime. The Agriculture Intensive (AI) class showed a decreased NDVI value from June until August, a period that is considered quite dry. Taller vegetation types (e.g., Closed Evergreen Tropical Forest (CETF)) maintained NDVI values above 0.8 from May to September. These higher NDVI values during the dry season could be related with the ability of vegetation to capture depth water.

The NDVI time series training profiles show a considerable similarity among a number of classes (**Figure 5**), which naturally influences the classification results. **Figure 6** presents the classified images for the reference year (2000) using the ED and DTW classifiers. The majority of Rondonia was classified in both images as CETF. Large differences between ED and DTW were obtained in the eastern region of Rondonia: the AI and Grass Savannah (GS) classes were dominant using the ED classifier, but when using the DTW, only the AI could be well distinguished. Some of the pixels that were classified as GS using the ED classifier were classified as Closed Shrublands (CS) using the DTW. This is because these two classes were composed predominantly of grasses and shrubs. The GS class is very dependent on water availability and is characterized by open canopy with herbaceous soil vegetation, whereas the CS class is characterized by shrubs, herbs, and grasses as dominant vegetation, and this vegetation can remain stable over time or can be affected by factors such as fire or browsing. In both cases, the GS and CS classes presented a similar profile in the main growing season (**Figure 5**), and they can present a double growth due to the herbaceous soil vegetation growth, animal grazing, human interference, or other adverse conditions such as fire.

In the western region of Rondonia another difference can be easily observed between the ED and DTW images: some of the pixels that were classified as CETF using the ED classifier were classified as Montane Forest – Dense Evergreen (MFDE) using the DTW. This can be explained because CETF and MFDE are two classes of large evergreen forest vegetation, presenting high NDVI values and a stable behavior in the main growth season.

Using the profiles from the year under evaluation

The classified images for 1999, 2000, and 2011 using the ED and DTW classifiers are presented in **Figure 7;** 1999 presented better results with the reference year (2000) than 2011. The percentage of pixels that suffered changes between 1999 and 2000 (49% for ED and 57% for DTW) was lower compared with changes that occurred between 2000 and 2011 (66% for ED and 72% for DTW). This can be explained because larger temporal distance will, in general, correspond to higher land cover variability, resulting in higher differences in the classification results.

Table 4 presents the fraction of pixels with different classification results in consecutive years. For both classifiers, the most similar years were found to be 1999 and 2000. However, the maximum change between consecutive years occurred in different dates for ED (2001–2002) and DTW (2006–2007). The average fraction of pixels that changed between consecutive years was about 70% for both classifiers considering 15 LCC.

Using the profiles from the year 2000

Figure 8 shows the classified images for 1999, 2000, and 2011 using the ED and DTW classifiers and the reference signatures from 2000. Using the ED, 46% and 47% of pixels

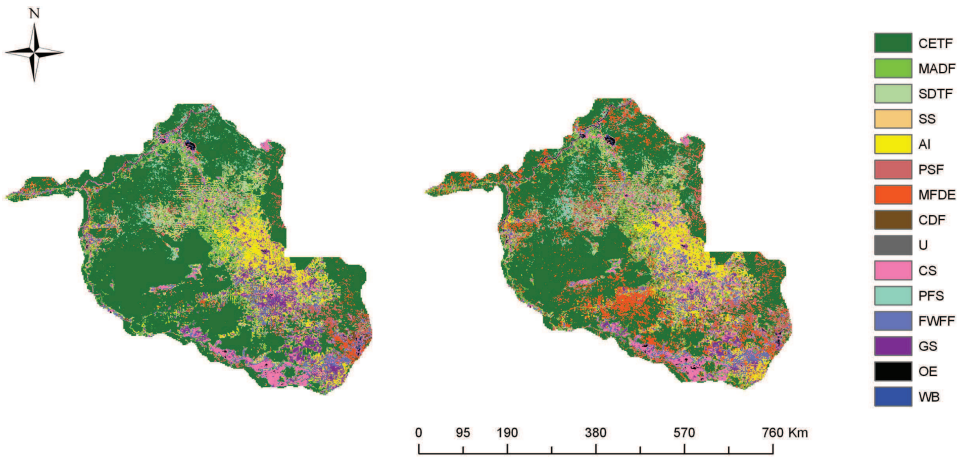


Figure 6. Classified images for the year 2000, using the Euclidean Distance (left) and Dynamic Time Warping (right) classifiers. (See **Table 2** for definitions of acronyms.)

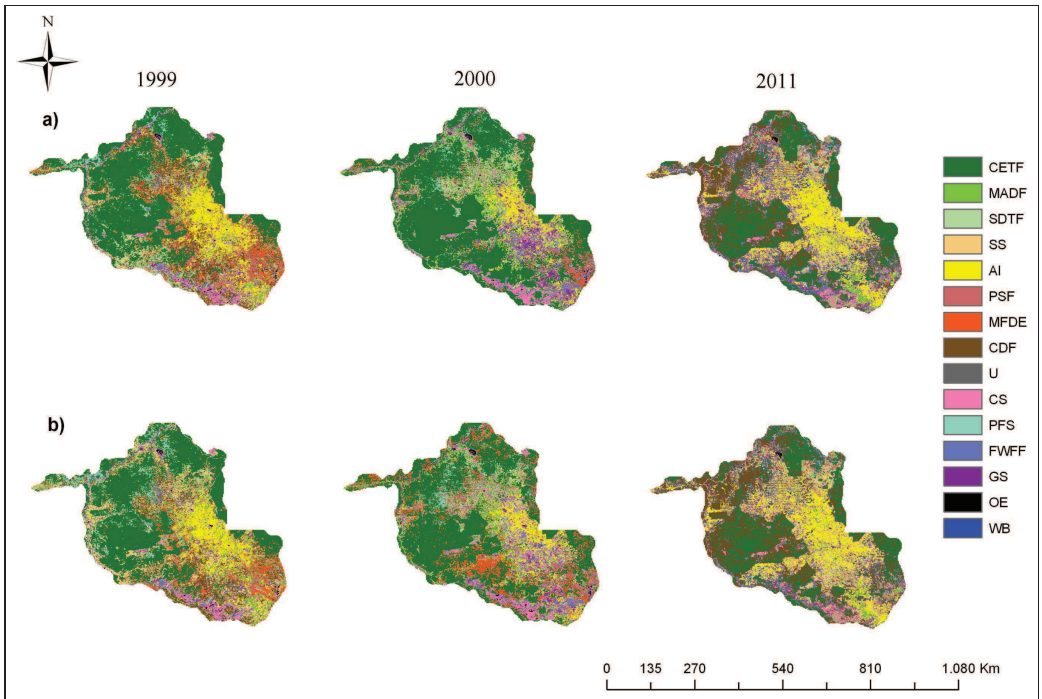


Figure 7. Classified images obtained for 1999, 2000, and 2011 using (a) Euclidean Distance and (b) Dynamic Time Warping as classifiers. (See **Table 2** for definitions of acronyms.)

Table 4. Fraction of pixels (%) that present different classification results between consecutive years.

	Profiles from the year under analysis		Profiles from 2000	
Years	ED classifier	DTW classifier	ED classifier	DTW classifier
1999–2000	49	57	46	50
2000–2001	71	75	70	66
2001–2002	78	73	64	69
2002–2003	75	74	53	63
2003–2004	73	71	45	63
2004–2005	74	69	38	54
2005–2006	73	75	57	56
2006–2007	74	77	63	62
2007–2008	66	68	67	64
2008–2009	71	72	62	56
2009–2010	74	68	46	47
2010–2011	60	62	39	45

Note: ED, Euclidean Distance; DTW, Dynamic Time Warping.

were found to change between 1999 and 2000 and between 2000 and 2011, respectively. Comparing this with the results obtained in the previous approach, the ED can reduce the fraction of pixels with changes between 1999 and 2000 by 3% and by 19% between 2000 and 2011. The DTW using profiles from 2000 also indicated less pixel changes (by 7% and 14%) between 1999 and 2000 and between 2000 and 2011.

Table 4 presents the fraction of pixels that have different classification results between consecutive years, using the reference signatures from 2000. Compared with the results obtained in the previous approach, the use of the 2000 profiles provided a higher correlation between images of two consecutive years and indicated a reduction on the average fraction of pixels with changes (16% for ED and 12% for DTW).

Effect of the NDVI time series similarity on the classification results

The two previous approaches indicated a considerable variability in LCC over the range of years tested. However, it is worth noting the large number of classes considered and the similarity between some of NDVI time series profiles used to train the LCC (**Figure 5**). To test the effect of the NDVI time series similarity on the classification results, the 15 LCC were reduced by joining the most similar classes. A hierarchical aggregation, based on the LCC similarity, was done using the ED and DTW measures (**Figure 9**). Through the dendogram analysis presented in **Figure 9**, it was possible to clearly identify four groups (A, B, C, and D). Group A contains four classes (2 (Mosaic Agriculture – Degraded Forest (MADF)), 3 (SemiDeciduous Transition Forest (SDTF)), 4 (Shrub Savannah (SS)) and 8 (Closed Deciduous Forest (CDF)), the group B contains four classes

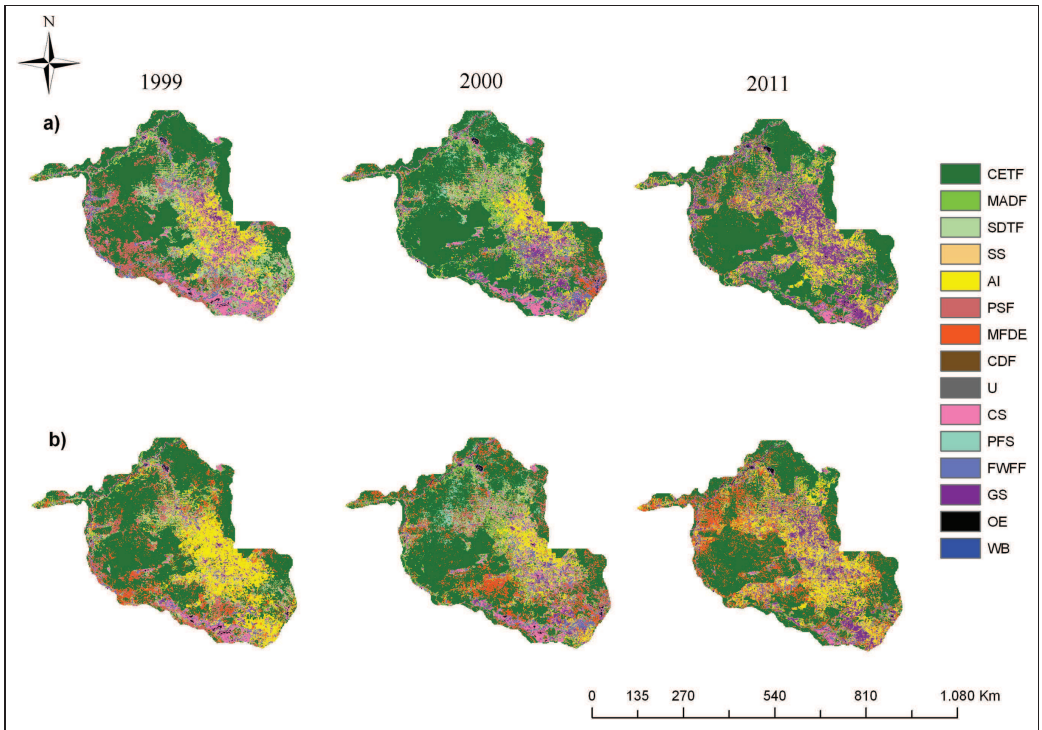


Figure 8. Classified images obtained for 1999, 2000, and 2011 using (a) Euclidean Distance and (b) Dynamic Time Warping as classifiers, and using the profiles of the reference year. (See **Table 2** for definitions of acronyms.)

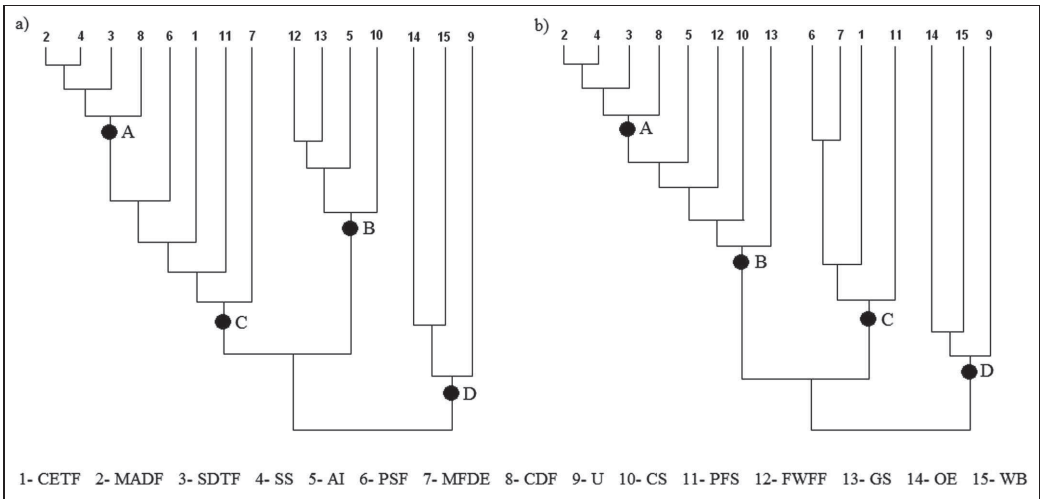


Figure 9. Dendograms for the hierarchical aggregation of the 15 land cover classes, based on the VGT data from 2000, using (a) the Euclidean Distance and (b) Dynamic Time Warping. (See **Table 2** for definitions of acronyms.)

(5 (AI), 10 (CS), 12 (Fresh Water Flooded Forest (FWFF)) and 13 (GS)), and the group C also four classes (1 (CETF), 6 (Permanent Swamp Forest (PSF)), 7 (MFDE) and 11 (Periodically Flooded Savannah (PFS))). Group D contains the three classes with lower NDVI (9 (Urban (U)), 14 (Ore Exploration (OE)) and 15 (Water Bodies (WB))).

Each of the classified images was reclassified using the approaches presented in the two previous sections, replacing each pixel class by the number of the enclosing group (from 1 to 4, corresponding to A–D). A pixel-by-pixel analysis was made to evaluate the land cover variability over the 13-year period (1999–2011). Each pixel was classified as

permanent, stable, change, stable in the first years, stable in the last years, and unspecified using the criteria presented in **Table 3** and **Figure 3**.

Table 5 presents a comparison between the ED classification results for four distinct pixels (P1, P2, P3, P4), using the 15 initial LCC and the four groups of classes. Using 15 classes, P1 and P2 presented a total of five distinct LCC over the 13 years, and P4 presented the lower land cover variability with a total of three LCC. Although a pixel can present higher land cover variability over the years, this does not imply a greater deviation between the land cover profiles. For example, P1 and P3 had an initial number of five and four distinct LCC, respectively, but using the aggregation groups the final number of classes obtained for both pixels were identical. This difference in LCC variability over the years influences the characterization of each pixel. This was the case for P1 and P2 which were classified as unspecified pixels using 15 LCC, but when using four classes they were classified as stable in the first years (P1) and permanent (P2) pixels.

Table 5. Classification results for four distinct pixels (P1, P2, P3, P4) from 1999 to 2011 using the Euclidean Distance.

	15 classes				4 classes			
Year	P1	P2	P3	P4	P1	P2	P3	P4
1999	11	1	1	1	3	3	3	3
2000	1	1	6	1	3	3	3	3
2001	11	7	11	11	3	3	3	3
2002	8	8	11	11	1	1	3	3
2003	7	7	11	1	3	3	3	3
2004	7	11	6	1	3	3	3	3
2005	6	7	1	6	3	3	3	3
2006	1	7	6	6	3	3	3	3
2007	1	1	1	1	3	3	3	3
2008	8	8	8	1	1	1	1	3
2009	6	6	8	6	3	3	1	3
2010	11	1	8	1	3	3	1	3
2011	8	1	8	1	1	3	1	3
Classes (no.)	5	5	4	3	2	2	2	1
Attribute*	6	6	5	5	4	1	3	1

*Attributes: 1, Permanent; 2, Stable; 3, Change; 4, Stable in the first years; 5, Stable in the last years; 6, Unspecified.

Figure 10 presents the classified images for the year 2000, using the four groups of classes. The land cover maps, obtained using ED and DTW classifiers, presented slight differences, instead of what happens when using 15 classes (**Figure 6**). Superimposing the ED and DTW land cover maps, the percentages of agreement obtained were 74.3% and 92.5% using 15 and four classes, respectively. In **Figure 10** it is possible to clearly identify the pixels with lower NDVI values (blue), such as water spots. The largest area of Rondonia was classified as group C (CETF, PSF, MFDE, PFS) using four classes, and as CETF using 15 LCC (**Figure 6**). These results demonstrate the consistency of the land cover maps before and after the aggregation of classes.

Table 6 presents the fraction of pixels classified in each attribute (defined in **Table 3** and **Figure 3**), using the approaches previously evaluated. The number of pixels classified as permanent over the 13-year period increased considerably with the aggregation of the 15 classes. Furthermore, the number of pixels unspecified was greatly reduced. The results presented in **Table 6** reinforce the idea that using the profiles from 2000 to classify all the years reduces the differences on the training profiles and, consequently, more similarity between the classification images over the years is achieved. Reducing the number of initial classes, aggregating the most similar, reduces the variability in the LCC over the years, providing more similar classification results between years and classifiers.

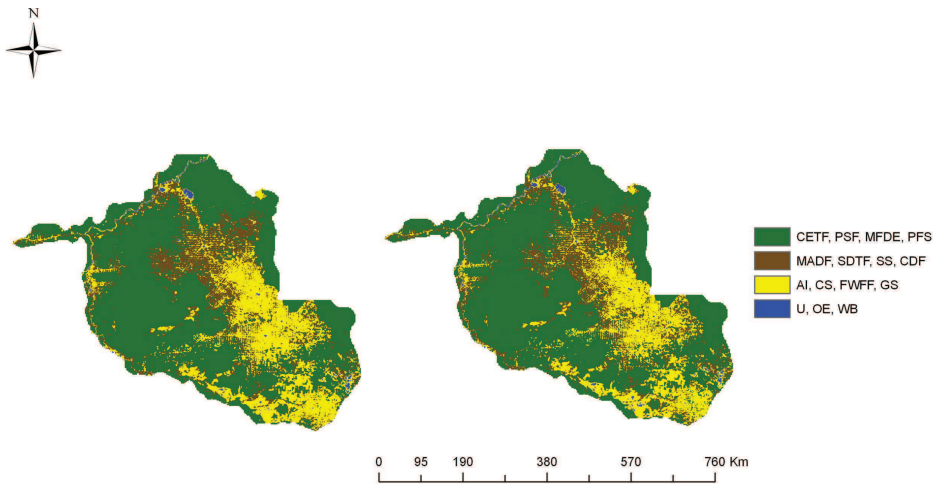


Figure 10. Classified images for the year 2000, using four groups of classes and using the Euclidean Distance (left) and Dynamic Time Warping (right) as classifiers. (See **Table 2** for definitions of acronyms.)

Evaluation of the classification results using the PRODES data

To evaluate the classification results obtained with the proposed methodology the PRODES data were used as a reference. **Figure 11** (left) presents the classified maps for Rondonia obtained for 2000 and 2011 using the PRODES data. In both maps, the main region of Rondonia was classified as forest and it is easy to identify the water-covered area, in particular the area in the upper right corner of the images. The deforestation has increased considerably between 2000 and 2011, occupying the biggest area of the central region of Rondonia.

To compare the ED and DTW land cover changes between 2000 and 2011 with the reference (PRODES) data, a reclassification was made aggregating the 15 initial LCC in three main classes: forest (enclosing the classes characterizing any type of forest vegetation (CETF, CDF, SDTF, PSF, MADF, MFDE, FWFF)), nonforest (AI, CS, GS, PFS, SS, U) and hydrography (WB, OE). The results of this reclassification, available in **Figure 11**, show coherent results between ED, DTW, and PRODES land cover maps. In both cases, the majority of Rondonia was classified as forest and the water spot in the upper right corner can be easily identified. For the new ED and DTW land cover maps, the percentage of pixels classified at each class was determined and compared with the PRODES data results (**Table 7**). The nonforest and deforestation classes of PRODES (reference) land cover maps were aggregated in a single class (nonforest). The results obtained based on the ED and DTW methodologies were consistent with those obtained from the reference data. The forest class occupies

Table 6. Fraction of pixels (%) classified per attribute from 1999 to 2011.

	15 classes				4 classes			
Attribute	ED	DTW	ED*	DTW*	ED	DTW	ED*	DTW*
1 – Permanent	1.51	0.67	18.97	9.81	36.3	37.18	68.80	65.18
2 – Stable	6.53	3.68	8.43	10.30	5.22	5.60	1.65	2.58
3 – Change	0.43	0.38	0.51	0.59	3.54	4.02	4.00	3.00
4 – Stable in the first years	12.69	9.03	13.27	13.17	29.1	28.13	9.51	11.08
5 – Stable in the last years	13.70	18.37	15.26	22.55	8.02	8.24	8.64	8.71
6 – Unspecified	65.14	67.88	43.56	43.59	17.7	16.83	7.39	9.45

*Profiles from 2000.
Note: ED, Euclidean Distance; DTW – Dynamic Time Warping.

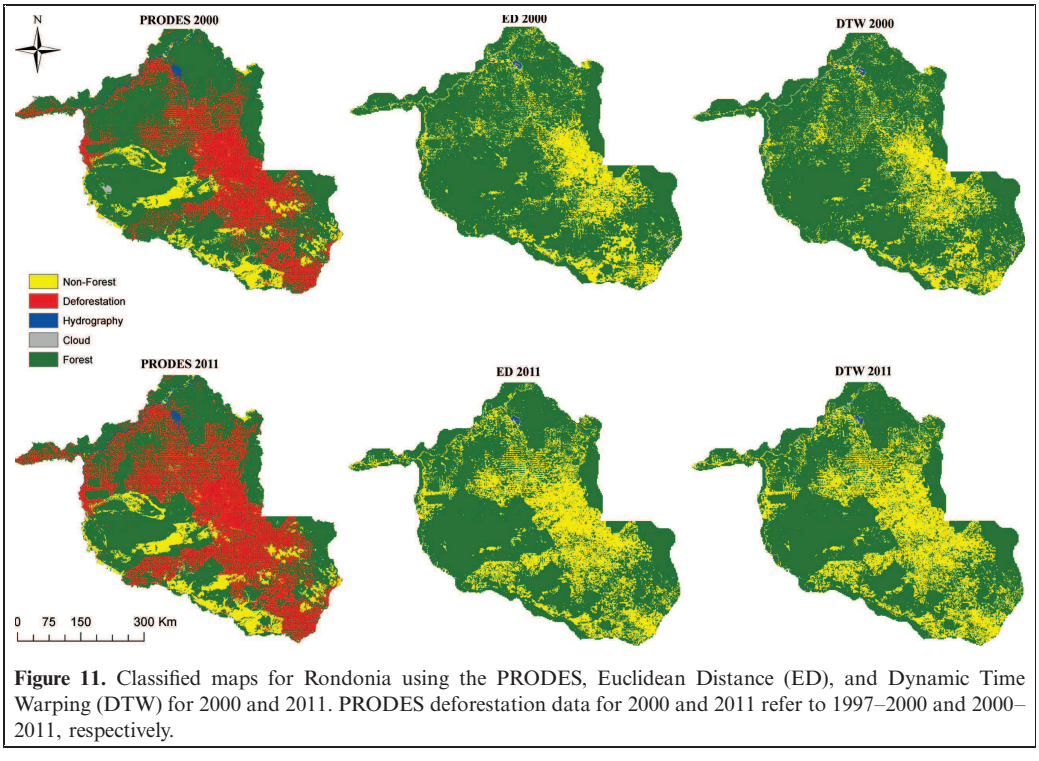


Figure 11. Classified maps for Rondonia using the PRODES, Euclidean Distance (ED), and Dynamic Time Warping (DTW) for 2000 and 2011. PRODES deforestation data for 2000 and 2011 refer to 1997–2000 and 2000–2011, respectively.

Table 7. Percentage of pixels classified at each class, using PRODES, ED, and DTW classifications for the years 2000 and 2011.

	Year	Forest (%)	Nonforest (%)	Hydrography (%)	Clouds (%)
PRODES	2000	62.96	36.09	0.77	0.18
	2011	53.01	46.15	0.78	0.07
Euclidean Distance (ED)	2000	77.76	21.83	0.41	—
	2011	69.36	30.47	0.18	—
	2011*	65.50	34.14	0.36	—
Dynamic Time Warping (DTW)	2000	77.99	21.46	0.54	—
	2011	67.71	32.03	0.25	—
	2011*	68.38	31.26	0.36	—

*Profiles from 2000.

the majority of Rondonia, always higher than 53% areal coverage. The reference data presented a 9.95% decrease in the forest area, due to the increase of deforestation in the Rondonia region between the years 2000 and 2011 (**Figure 11**, left). The same conclusion was verified using the ED and DTW classifications, where a decrease (increase) of around 10% was verified in the forest (nonforest) class (**Figure 11**). The variability between hydrography can be due to the fact that the urban class (considered as nonforest) was sometimes classified as hydrography due to the low NDVI values and similar profile with WB and OE classes.

Conclusions

A new approach was proposed for land cover classification and land cover map updating based exclusively on temporal series data acquired by EOS. The approach was illustrated with a practical test using SPOT VGT data from 1999 to 2011 in the Brazilian Amazon. The GLC2000 was used as a reference to select 15 distinct LCC in the SPOT VGT images of Rondonia.

Maps of land cover changes for 1999–2011 were obtained using two classifiers: one that represents a time-rigid measure (ED) and another that represents a time-flexible measure (DTW). The similarity between some NDVI profiles had influenced the classification results, leading to some differences between the ED and DTW land cover maps, particularly in the eastern and western regions of Rondonia.

The noise due to the cloud contamination (especially from November to February) and the characteristics of the study area makes it difficult to classify 15 LCC in the Amazon region. Thus, the 15 initial LCC were gradually aggregated, joining the most similar, until only four final broad classes remained. The land cover variability over time was evaluated by making a pixel-by-pixel analysis. Each pixel was classified as permanent, stable, change, stable in the first years, stable in the last years, and unspecified using a criteria based on the number of identical classes from 1999 to 2011. The maps of land cover changes obtained using 15 and four LCC presented coherent results. As expected, the reduction of the number of classes allowed higher similarity over the years.

The PRODES data were used to evaluate the results of the ED and DTW classifications for the years 2000 and 2011. The ED and DTW land cover maps showed consistent results comparing with PRODES data, which proves that this approach is capable to produce accurate land cover maps for Rondonia, using exclusively temporal series data.

Acknowledgments

The authors would like to thank Flemish Institute for Technological Research (VITO) for the SPOT VEGETATION images, and also to the Global Vegetation Monitoring Unit of the European Commission Joint Research Centre, for providing access to the GLC2000. A. Rodrigues would like to thank to Fundação para a Ciência e a Tecnologia (FCT) for the Doctoral Grant (SFRH/BD/62189/2009). D. Furlan would like to thank to the Santander Scholarship Program for International Mobility and to the Research Foundation of the State of São Paulo – FAPESP (2010/02228-0), who supported the research period spent at the University of Porto.

References

Agrawal, R., Faloutsos, C., and Swami, A. 1993. Efficient similarity search in sequence databases. *Lectures Notes in Computer Science*, Vol. 730, pp. 69–84.

Agrawal, R., Lin, K., Sawhney, H.S., and Shim, K. 1995. Fast similarity search in the presence of noise, scaling, and translation in times-series databases. *Proceedings of the 21st International Conference on Very Large Data Bases*, pp. 490–501.

Alves, D.S., Morton, D.C., Batistella, M., Roberts, D.A., and Souza Jr., C. 2009. The changing rates and patterns of deforestation and land use in Brazilian Amazonia. *Geophysical Monograph Series*, Vol. 186, pp. 11–23.

Alves, D.S., Pereira, J.L.G., De Sousa, C.L., Soares, J.V., and Yamaguchi, F. 1999. Characterizing landscape changes in central Rondonia using Landsat TM imagery. *International Journal of Remote Sensing*, Vol. 20, No. 14, pp. 2877–2882. doi: 10.1080/014311699211859.

ANA - Agência Nacional De Águas (Brasil). Sistemas de Informações Hidrológicas: http://hidroweb.ana.gov.br/HidroWeb [Accessed 7 March 2012].

Asner, G.P., Keller, M., Lentini, M., Merry, F., and Souza Jr., C. 2009. Selective logging and its relation to deforestation. *Geophysical Monograph Series*, Vol. 186, pp. 25–42.

Ballester, M.V.R., Victoria, D.C., Krusche, A.V., Coburn, R., Victoria, R.L., Richey, J.E., Logsdon, M.G., Mayorga, E., and Matricardi, E. 2003. A remote sensing/GIS-based physical template to understand the biogeochemistry of the Ji-Parana river basin (Western Amazonia). *Remote Sensing of Environment*, Vol. 87, No. 4, pp. 429–445. doi: 10.1016/j.rse.2002.10.001.

Berndt, D.J., and Clifford, J. 1994. Using Dynamic Time Warping to find patterns in time series. In *Workshop on Knowledge Discovery in Databases*, Seattle, USA, pp. 359–370.

Braswell, B.H., Hagen, S.C., Frolking, S.E., and Salas, W.A. 2003. A multivariable approach for mapping sub-pixel land cover distributions using MISR and MODIS: Application in the Brazilian Amazon region. *Remote Sensing of Environment*, Vol. 87, pp. 243–256. doi: 10.1016/j.rse.2003.06.002.

Brown, J.C., Jepson, W.E., Kastens, J.H., Wardlow, B.D., Lomas, J.M., and Price, K.P. 2007. Multitemporal, moderate-spatial-resolution remote sensing of modern agricultural production and land modification in the Brazilian Amazon. *GIScience & Remote Sensing*, Vol. 14, No. 2, pp. 117–148. doi: 10.2747/1548-1603.44.2.117.

Carreiras, J.M.B., Pereira, J.M.C., Shimabukuro, Y.E., and Stroppiana, D. 2003. Evaluation of compositing algorithms over the Brazilian Amazon using SPOT-4 Vegetation data. *International Journal of Remote Sensing*, Vol. 24, No. 17, pp. 3427–3440. doi: 10.1080/0143116021000021251.

Chu, S., Keogh, E., Hart, D., and Pazzani, M. 2002. Iterative deepening Dynamic Time Warping for time series. In *Proceedings of the 2nd SIAM International Conference on Data Mining*, Maebashi City, Japan.

Dale, V.H., O'Neill, R., Southworth, F., and Pedlowsky, M. 1994. Modeling effects of land management in the Brazilian Amazonian settlement of Rondonia. *Conservation Biology*, Vol. 8, No. 1, pp. 196–206. doi: 10.1046/j.1523-1739.1994.08010196.x.

Da Rocha, H.R, Goulden, M.L., Miller, S.D., Menton, M.C., Pinto, L.D.V.O, De Freitas, H.C., and Figueira, A.M.S. 2004. Seasonality of water and heat fluxes over a tropical forest in eastern Amazon. *Ecological Application*, Vol. 14, No. 4, Supplement, pp. S22–S32. doi: 10.1890/02-6001.

Das, G., Lin, K., Mannila, H., Renganathan, G., and Smyth, P. 1998. Rule discovery from time series. *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, pp. 16–22.

Debregeas, A., and Hebrail, G. 1998. Interactive interpretation of Kohonen maps applied to curves. *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, pp. 179–183.

Eva, H.D., Miranda, E.E., Di Bella, C.M., Gond, V., et al. 2002. Vegetation map of South America. *Joint Research Centre – EUR20159*, Office for Publications of the European communities, Luxembourg.

Eva, H.D., Belward, A.S., Miranda, E.E., et al. 2004. A land cover map of South America. *Global Change Biology*, Vol. 10, pp. 31–744. doi: 10.1111/j.1529-8817.2003.00774.x.

Eva, H.D., Gallego, J., Strahler, et al. 2006. Validation of the global land cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, No. 7, pp. 1728–1739. doi: 10.1109/TGRS.2006.864370.

Faloutsos, C., Ranganathan, M., and Manolopoulos, Y. 1994. Fast subsequence matching in time-series databases. *ACM SIGMOD Record*, Vol. 23, No. 2, pp. 419–429. doi: 10.1145/191843.191925.

Fujisaka, S., Bell, W., Thomas, N., Hurtado, L., and Crawford, E. 1996. Slash-and-burn agriculture, conversion to pasture, and deforestation in two Brazilian Amazon colonies. *Agriculture, Ecosystems & Environment*, Vol. 59, No. 1–2, pp. 115–130. doi: 10.1016/0167-8809(96)01015-8.

Gonsamo, A., and Chen, J.M. 2011. Evaluation of the GLC2000 and NALC2005 land cover products for LAI retrieval over Canada. *Canadian Journal of Remote Sensing*, Vol. 37, No. 3, pp. 302–313. doi: 10.5589/m11-039.

Houghton, P.A., Skole, D.L., Nobre, C.A., Hackler, J.L., Lawrence, K.T., and Chomentowski, W.H. 2000. Annual fluxes of carbon from deforestation and regrowth in Brazilian Amazon. *Nature*, Vol. 403, pp. 301–304. doi: 10.1038/35002062.

Keogh, E.J., Chakrabarti, K., Pazzani, M.J., and Mehrotra, S. 2001. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, Vol. 3, No. 3, pp. 263–286. doi: 10.1007/PL00011669.

IBGE (Instituto Brasileiro de Geografia e Estatistica) 2013. Available at http://ibge.gov.br/home/ [Accessed 16 February 2013].

INPE 2000. Monitoring of the Brazilian Amazon forest by satellite 1999–2000. Available at http://mtc-m18.sid.inpe.br/col/dpi.inpe.br/lise/2001/05.16.09.55/doc/html/capa.htm [Accessed 16 February 2013].

Li, G., Lu, D., Moran, E., and Hetrick, S. 2012. Land-cover classification in a moist tropical region of Brazil with Landsat TM imagery. *International Journal of Remote Sensing*, Vol. 32, No. 23, pp. 8207–8230. doi: 10.1080/01431161.2010.532831.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., and Merchant, J.W. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, Vol. 21, pp. 1303–1330. doi: 10.1080/014311600210191.

Malhi, Y., Pegoraro, E., Nobre, A.D., Pereira, M.G.P., Grace, J., Culf, A.D., and Clement, R. 2002. Energy and water dynamics of a central Amazonian rain forest. *Journal of Geophysical Research: Atmospheres*, Vol. 107, No. D20, pp. LBA 45-1–LBA 45-17. doi: 10.1029/2001JD000623.

Moran, E.F. 1993. Deforestation and land use in the Brazilian Amazon. *Human Ecology*, Vol. 21, No. 1, pp. 1–21. doi: 10.1007/BF00890069.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. 2007. Numerical Recipes: The Art of Scientific Computing, 3rd Edition. Cambridge University Press, pp. 766–772.

Pedlowski, M.A., Dale, V.H., Matricardi, E.A.T., and Filho, E.P.S. 1997. Patterns and impacts of deforestation in Rondonia, Brazil. *Landscape and Urban Planning*, Vol. 38, No. 3–4, pp. 149–157. doi: 10.1016/S0169-2046(97)00030-3.

PRODES (Programa de Cálculo do Desflorestamento da Amazônia) 2013. Available at http://www.obt.inpe.br/prodes/index.php [Accessed 16 January 2013].

Reed, G.F., Lynn, F., and Meade, B.D. 2002. Use of coefficient of variation in assessing variability of quantitative assays. *Clinical and Vaccine Immunology*, Vol. 9, No. 6, pp. 1235–1239. doi: 10.1128/CDLI.9.6.1235-1239.2002.

Rodrigues, A., Marçal, A.R.S., and Cunha, M. 2013. Identification of land cover changes on a continental scale using NDVI time-series from SPOT VEGETATION. *International Journal of Remote Sensing*, Vol. 34, No. 22, pp. 8028-8050.

Rouse, J.W., Haas, R.H., Schell, J.A., and Deering, D.W. 1973, Monitoring vegetation systems in Great Plains with ERST. *Proceedings of the 3rd ERTS Symposium, NASA SP-351*, 10–14 December 1973, Washington, DC (US Government printing office), pp. 309–317.

VITO (Free SPOT VEGETATION products) 2012. Available at http://free.vgt.vito.be/ [Accessed 10 August 2012].

Wardlow, B.D., and Egbert, S.L. 2008. Large-area crop mapping using time-series MODIS 250m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sensing of Environment*, Vol. 112, pp. 1096–1116. doi: 10.1016/j.rse.2007.07.019.

CONCLUSIONS

Conclusions

The work carried out focused on the development of the PhenoSat tool, and on the application of temporal series of VI for land cover classification. There are other potential applications of phenological parameter extraction from EOS data, such as crop value and quality evaluation, but they were not pursued in this thesis.

The PhenoSat software, developed to extract phenological information from EOS temporal vegetation index data, proved to be capable of obtaining accurate and consistent results when compared with field observations.

Currently, the access to low spatial resolution EOS data is easy and with low or no cost, making these data prevalent in the development of tools to process and extract phenological metrics. However, the presence of noise in the EOS temporal series is a frequent and relevant aspect that requires processing techniques to minimize its effects. PhenoSat uses six fitting methods to smooth the vegetation satellite data: Savitzky-Golay (SG), Cubic Smoothing Splines (CSS), Polynomial Curve Fitting (PCF), Gaussian models (GM), Fourier Series (FS) and Piecewise-Logistic (PL). Four of these methods (CSS, PCF, GM and FS) depend on a smoothing parameter, which determines how closely the smoothing data adheres to the original data. All six methods proved capable of reducing the noise, without suppressing the natural variations of the vegetation. These smoothing processes permit a precise and accurate subsequent data analyses (more detailed information in appendices 2 and 3).

Comparing with other software packages available, PhenoSat presents two major advantages: (i) the detection and extraction of phenological information for annual regrowth and (ii) the selection of an in-season region of interest.

In all the experiments carried out, PhenoSat was capable of identifying the double growth season occurrences. It obtained accurate phenological information for the beginning and maximum of the regrowth period, with the differences between estimated and field observed measures not higher than 10-days.

The selection of a region of interest allows the reduction of the processing time in addition to a better fitting and more accurate phenological measures (for more details see appendix 3, section 4.2). This feature proved to be very valuable to monitor vineyard and other crop with discontinuous canopy, like forestry or deciduous fruit trees. The in-season

region of interest should be defined using the crops behaviour in the field at normal conditions (regular atmospheric conditions and no human interventions that can affect the normal growth period).

The algorithm based on growth rate changes to extract the phenological information, avoid the use of thresholds or empirical constants, making PhenoSat a flexible tool to be applied on different crops and vegetation index data from different data sources.

Since PhenoSat was released online (http://www.fc.up.pt/PhenoSat), on July 2011, important feedback was received which led to successive improvements (e.g. appendices 2 and 3). One of that improvements is related with the PhenoSat inputs. Initially, PhenoSat received as input the temporal VI dataset values, requiring the previous extraction of these values from the EOS images. Depending on the number of years/datasets in analysis, this can be a time-consuming process. To solve this problem, PhenoSat is now also able to receive the original VI images as input, which can be used entirely or just on a specific region.

Another improvement is the creation of digital phenological maps. Three stages, considered to represent the principal vegetation growth cycle parameters, were chosen: the Maximum Vegetation Development (MVD), the start of season and the end of season. MVD indicates the maximum annual vegetation value achieved, and is generally related to a time during the growing season when biomass is high and growth is rapid; the start and end of season are the main phenological indicators, and they are always considered in the studies that have investigated vegetation dynamics using information from date phenometrics.

Another task that can be done in the future, to promote the PhenoSat application, is to rewrite it in a more flexible programming language, like C or C#, which would facilitate the distribution and avoid licence costs and Matlab´s installation constraints (such as version and toolboxes).

The work developed for land cover change detection culminated in a classification methodology for updating land cover maps on large scales. The proposed method uses exclusively vegetation time-series data acquired by EOS, and is based on a hierarchical aggregation of the land cover classes. The purpose of the hierarchical aggregation is to evaluate the two proposed methods: New Signature (NS) and Preserve Signature (PS). As the signature of the new class (obtained by joining the two most similar ones) is modified in NS and not in PS, the classification results will be distinct. With this procedure

it is possible to evaluate if the differences between NS and PS are significant and, consequently, if there are advantages on using one method instead of the other. For both PS and NS methods, improved classification results were obtained as the aggregation level increased. However, NS tends to be slightly worse due to the modification of spectral signatures at each level.

In general, the classified results obtained using the proposed methodology were consistent with those obtained from reference data, proving that this approach is able to produce accurate land cover maps using exclusively temporal series of EOS data. For both studies carried out (on the African Continent and on the Brazilian Amazon) it was verified that the aggregation of the most similar land cover classes allowed the reduction of the variability, thus improving the classification accuracies.

The choice of the distance measure to perform the hierarchical aggregation plays a significant role in the quality of the classification algorithm. Four similarity measures were tested: Euclidean Distance (ED), Cosine (Cos), Jaccard (Jac), and Dynamic Time Warping (DTW). ED is the most straightforward similarity measure for time-series. However it requires that the two time-series in comparison have the same length. Besides that, ED, Cos and Jac do not consider shifts or distortions in time, which is a relevant aspect when comparing temporal vegetation data. Thus, DTW is used to solve this limitation. It is characterized by a non-linear mapping between two sequences, where the distance between them is minimized. The calculation of DTW was more demanding, but it proved to be more efficient when analysing land cover changes over the years. In addition to DTW, new time-flexible measures can be further implemented and tested, as for example the Longest Common Subsequence, the Edit Distance with Real Penalty and the Edit Distance on Real Sequence. These similarity measures address the problems related with local shifting and presence of noise.

In the proposed temporal-based classification approach, the user indicates the initial land cover classes and the classification is done without considering *a priori* any inter-relationships between the initial classes. The hierarchical arrangement, and the distance measure used to form a new class, is very often in conflict with the ability to define a clear boundary between two classes. For example, in chapter III, section 2, figure 2, it is possible to see a higher similarity between the profiles of the BR (bare rock), SDD (sandy desert and dunes) and SH (salt hardpans) classes. Although the similarity of these classes have resulted in their aggregation (chapter III, section 3, figures 8, 9 and 10), this depends on the similarity measure being used.

To ensure the aggregation of similar classes during the hierarchical aggregation, a new approach, consisting in incorporating some pre-defined hierarchical relationship at the beginning of the classification process, can be used. This approach can serve the distinct requirements of different end users. For example, a forester and a rural planner can have different interests: the first may be interested in preserving the information associated with the classes 'dense forests' and 'sparse forests' and in aggregating all remaining classes at a higher hierarchical level; whereas the last may be interested in separate 'cultivated area' and 'intensive cultivated area'. Thus, in a classification process it is important to understand what is the purpose of the analysis to ensure that the proper and most useful classes are identified.

Overall, both the PhenoSat tool and the classification methodologies developed have great potential for high level information extraction from EOS data.

Some new research lines were identified to follow the previous work: (1) the comparison of EVI and NDVI phenological measures for the same crop and region; (2) the combination of NDVI from AVHRR and other sensors to obtain a long-term NDVI time-series; and (3) the determination of indicators that allow to evaluate vegetation stresses as drought.

REFERENCES

References

Aban, J.E.L., R. Tsolmon, and R. Tateishi (2002). Linear mixing model based on optimization method for land cover mapping using Landsat ETM+ and SPOT-HRV data. *Proceedings of the 2002 IEEE International Geoscience and Remote Sensing Symposium, 6,* 3483-3485

Alberte, F. (1994). A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters. *Remote Sensing of Environment, 48,* 220-230

Allen, W.A., H.W. Gausman, A.J. Richardson, and J.R. Thomas (1969). Interaction of isotropic light with a compact plant leaf. *Journal of the Optical Society of America, 59,* 1376-1379

Allen, W.A., H.W. Gausman, and A.J. Richardson (1973). Willstater-stoll theory of leaf reflectance evaluated by ray tracing. *Applied Optics, 12,* 2448-2453

Atzberger, C., and F. Rembold (2009). Estimation of inter-annual winter crop area variation and spatial distribution with low resolution NDVI data by using neural networks trained on high resolution images. *Proceedings of SPIE, Remote Sensing for Agriculture, Ecosystems and Hydrology XI,* 7472.

Atzberger, C., and P.H.C. Eilers (2010). A smoothed 1-km resolution NDVI time series (1998-2008) for vegetation studies in South America. *International Journal of Digital Earth, 4*, 365-386

Badeck, F.W., A. Bondeau, K. Bottcher, D. Doktor, W. Lucht, J. Schaber, and S. Sitch (2004). Responses of Spring phenology to climate change. *New Phytologist, 162,* 295-309

Badhwar, G.D. (1984). Use of Landsat-derived profile features for spring small-grains classifications. *International Journal of Remote Sensing, 5,* 783-797

Basso, B., D. Cammarano, and P. De Vita (2004). Remotely sensed vegetation indices: theory and applications for crop management. *Italian Journal of Agrometeorology*, *1*, 36-53

Batjes, N.H., and W.G. Sombroek (1997). Possibilities for carbon sequestration in tropical and subtropical soils. *Global Change Biology, 3,* 161-173

Beck, P., C. Atzberger, K. Hogda, B. Johansen, and A. Skidmore (2006). Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. *Remote Sensing of Environment, 100*, 321-334

Bin, T., J. Morisette, R. Wolf, W. Esaias, G. Feng, G. Ederer, J. Nightingale, J.E. Nockeson, P. Ma, and J. Pedeley (2008). Vegetation phenology metrics derived from temporally smoothed and gap-filled MODIS data. *Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium, 3,* 593-596

Boegh, E., H. Soegaard, N. Broge, C.B. Hasager, N.O. Jensen, K. Schelde, and A. Thomsen (2002). Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sensing of Environment, 81,* 179-193

Bouman, B.A.M. (1995). Crop modeling and remote sensing for yield prediction. *Netherlands Journal of Agricultural Science, 43*, 143-161

Bradley, B., R. Jacob, J. Hermance, and J. Mustard (2007). A curve fitting procedure to derive inter-annual phenologies from time-series of noisy satellite NDVI data. *Remote Sensing of Environment, 106,* 137-145

Bradley, B.A., and J.F. Mustard (2008). Comparison of phenology trends by land cover class: a case study in the Great Basin, USA. *Global Change Biology, 14,* 334-346

Campbell, J.B. (2002). Introduction to Remote Sensing*. New York London: the Guilford Press*

Carreiras, J.M.B., J.M.C. Pereira, Y.E. Shimabukuro, and D. Stroppiana (2003). Evaluation of compositing algorithms over the Brazilian Amazon using SPOT-4 Vegetation data. *International Journal of Remote Sensing, 24,* 3427-3440

Chen, J., P. Jonsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment, 91*, 332-334

Chuine, I. (2000). A unified model for budburst of trees. *Journal of Theoretical Biology, 707,* 337-347

Cihlar, J. (2000). Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing, 21,* 1093-1114

Cihlar, J., D. Manak, and M. D'iorio (1994). Evaluation of compositing algorithms for AVHRR data over land. *IEEE Transactions on Geoscience and Remote Sensing, 32*, 427-437

Cihlar, J., H. Ly, Z.Q. Li, J. Chen, H. Pokrant, and F.T. Huang (1997). Multitemporal, multichannel AVHRR data sets for land biosphere studies – Artifacts and corrections. *Remote Sensing of Environment, 60,* 35-57

Colditz, R.R., C. Conrad, T. Wehrmann, M. Schmidt, and S. Dech (2008). TiSeG: A flexible software tool for time-series generation of MODIS data utilizing the quality assessment science data set. *IEEE Transactions on Geoscience and Remote Sensing, 46*, 3296-3308

Cunha, M., Marçal. A.R.S., and L. Silva (2010). Very early prediction of wine yield based on satellite data from VEGETATION. *International Journal of Remote Sensing, 31*, 3125-3142

de Beurs, K.M., and G.M. Henebry (2004). Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan. *Remote Sensing of Environment, 89,* 497-509

de Beurs, K.M., and G.M. Henebry (2010). Spatio-temporal statistical methods for modelling land surface phenology. *I.L. Hudson and M.R. Keatley (Eds.), Phenological Research: Methods for Environmental and Climate Change Analysis*. Springer, 177-208

Eidenshink, J.C. (1992). The 1990 Conterminous U.S. AVHRR data set. *Photogrammetric Engineering and Remote Sensing, 74*, 609-620

Fisher, J.I., J.F. Mustard, and M.A. Vadeboncoeur (2006). Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment, 100*, 265-279

Fontana, F., C. Rixen, T. Jonas, G. Aberegg, and S. Wunderle (2008). Alpine grassland phenology as seen in AVHRR, VEGETATION and MODIS NDVI time series – A comparison with in situ measurements. *Sensors, 8*, 2833-2853

Gao, F., J.T. Morisette, R.E. Wolfe, G. Ederer, J. Pedelty, E. Masuoka, R. Myneni, B. Tan, and J. Nightingale (2008). An algortithm to produce temporally and spatially continuous MODIS-LAI time series. *IEEE Geoscience and Remote Sensing Letters, 5,* 60-64

Gausman, H.W., W.A. Allen, V.I. Myers, and R. Cardenas (1969). Reflectance and internal structure of cotton leaves *Gossypium hirsutum L.*. *Agronomy Journal, 61*, 374-376

Gausman, H.W., and W.A. Allen (1973). Optical Parameters of Leaves of 30 Plant Species. *Plant Physiology, 52*, 57-62

Gausman, H.W., and W.G. Hart (1974). Reflectance of sooty mold fungus on citrus leaves over the 2.5 to 40-micrometer wavelenght interval. *Journal of Economic Entomology, 67*, 479-480

Gausman, H.W., W.A. Allen, and D.E. Escobar (1974). Refractive Index of Plant Cell Walls. *Applied Optics, 13*, 109-111

Gitelson, A.A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology, 161,* 165-173

Gitelson, A.A., and Y.J. Kaufman (1998). MODIS NDVI optimization to fit the AVHRR data series-spectral considerations. *Remote Sensing of Environment, 66*, 343-350

Gobron, N., B. Pinty, M. Verstraete, and J. Widlowski (2000). Development of spectral indices optimized for the VEGETATION instrument. *Proceedings of VEGETATION2000*, 275-280

Hansen, M.C., and R.S. DeFries (2004). Detecting long-term global forest change using continuous fields of tree-cover maps from 8-km Advanced Very High Resolution Radiometer (AVHRR) data for the years 1982-99. *Ecosystems, 7,* 695-716

Hansen, M.C., R.S. DeFries, J.R.G. Townshend, and R.A. Sohlberg (2000). Global land cover classification at 1km spartial resolution using a classification tree approach. *International Journal of Remote Sensing, 21*, 1331-1364

Helbig, N., B. Vogel, H. Vogel, and F. Fiedler (2004). Numerical modelling of pollen dispersion on the regional scale. *Aerobiologia, 3,* 3-19

Hermance, J.F., R.W. Jacob, B.A. Bradley, and J. Mustard (2007). Extracting phenological signals from multiyear AVHRR NDVI time series: Framework for applying high-order annual splines with roughness damping. *IEEE Transactions on Geoscience and Remote Sensing, 45*, 3264-3276

Hird, J.N., and G.J. McDermid (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment, 113*, 248-258

Holben, B.N. (2007). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing, 7*, 1417-1434

Houghton, R.A. (2000). Interannual variability in the global carbon cycle. *Journal of Geophysical Research: Atmospheres, 105,* 20121-20130

Hubert, L.F., and O. Berg (1955). A rocket portrait of a tropical storm. *Monthly Weather Review, 83,* 119-124

Hudson, I.L., S.W. Kim, and M.R. Keatley (2009). Climatic influences on the flowering phenology of four Eucalypts: a GAMLSS approach. *I.L. Hudson and M.R. Keatley (Eds.), Phenological Research: Methods for Environmental and Climate Change Analysis*. Springer, 209-228

Huete, A.R. (1988). A soil adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25,* 295-309

Huete, A.R., and R.D. Jackson (1988). Soil and atmosphere influences on the spectra of partial canopies. *Remote Sensing of Environment, 25*, 89-105

Huete, A., H. Liu, K. Batchily, and W. van Leeuwen (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sening of Environment, 59*, 440-451

Huete, A., K. Didan, E.P. Rodriguez, X. Gao, and L.G. Ferreira (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environemnt, 83*,195-213

Huete, A., K. Didan, Y.E. Shimabukuro, P. Ratana, C.R. Saleska, L.R. Hutyra, W. Yang, R.R. Nemani, and R. Myneni (2006). Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters, 33*, L06405

Jiang, Z., A.R. Huete, K. Didan, and T. Miura (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment, 112,* 3833-3845

Jonsson, P., and L. Eklundh (2002). Seasonality extraction by function-fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing, 40*, 1824-1832

Jonsson, P., and L. Eklundh (2004). TIMESAT – a program for analyzing time-series of satellite sensor data. *Computers and Geoscience, 18*, 833-845

Julien, Y., and J.A. Sobrino (2009). Global land surfasse phenology trends from GIMMS database. *International Journal of Remote Sensing, 30,* 3495-3513

Justice, C.O., E. Vermote, J.R. Townshend, R. Defries, D.P. Roy, D.K. Hall, V.V. Salomonson, J.L. Privette, G. Riggs, A. Strahler, W. Lucht, R.P. Myneni, Y. Knyazikhin, S.W. Running, R.R. Nemani, Z. Wan, A.R. Huete, W. van Leeuwen, R.E. Wolfe, L. Giglio, J.P. Muller, P. Lewis, and M.J. Barnsley (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing, 36*, 1228-1249

Justice, A., A. Belward, J. Morisette, P. Lewis, J. Privete, and F. Baret (2000). Developments in the 'validation' of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing, 21*, 3383-3390

Justice, C.O., J.R. Townshend, E. Vermote, E. Masuoka, R.E. Wolfe, N. Saleous, D.P. Roy, and J. T. Morisette (2002). An overview of MODIS land data processing and product status. *Remote Sensing of Environment, 83*, 3-15.

Kogan, F.N. (1995). Application of vegetation index and brightness temperature for drought detection. *Advanced Space Research, 15,* 91-100

Kokalj, Ž., and K. Oštir (2007). Land cover mapping using Landsat satellite image classification in the classical Karst-Kras Region. *Acta Carsologica, 36*, 433-440

Langley, S.K., H.M. Cheshire, and K.S. Humes (2001). A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland. *Journal of Arid Environments, 49*, 401-411

Li, Z., and M. Kafatos (2000). Interannual variability of vegetation in United States and its relation to El Nino/Southern Oscillation. *Remote Sensing of Environment, 71*, 239-247

Lloyd, D. (1990). A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. *International Journal of Remote Sensing, 11*, 2269-2279

Loveland, T.R., B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L. Yang, and J.W. Merchant (2000). Development of a global land cover characteristics database and IGBP DISCover from 1km AVHRR data. *International Journal of Remote Sensing, 21,*1303-1330

Lovell, J.L., and R.D. Graetz (2001). Filetring pathfinder AVHRR land NDVI data for Australia. *International Journal of Remote Sensing, 22*, 2649-2654

Lunetta, R.S., J.F. Knight, J. Ediriwickrema, J.G. Lyon, and L.D. Worthy (2006). Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Senising of Environment, 105,* 142-254

Ma, M., and F. Veroustraete (2006). Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. *Advances in Space Research, 37,* 835-840

Mack, P.E. (1990). Viewing the Earth: The Social Construction of the Landsat Satellite System (Inside Technology). *MIT Press*, Cambridge.

Marçal, A.R.S., J.S. Borges, J.A. Gomes, and J.F. Pinto da Costa (2005). Land Cover Update by Supervised Classification of Segmented ASTER Images. *International Journal of Remote Sensing, 26,* 1347–1362

Mayaux, P., H. Eva, A. Brink, F. Achard, and A. Belward (2008). Remote sensing of land-cover and land-use dynamics. *E. Chuvieco (Ed.), Earth Observation of Global Change.* Springer, 85-108

McCloy, K.R., and W. Lucht (2004). Comparative evaluation of seasonal patterns in long time series of satellite image data and simulations of global vegetation model. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 140-153

McKellip, R., D. Prados, R. Ryan, K. Ross, J. Spruce, G. Gasser, and R. Greer (2008). Remote-sensing time series analysis, a vegetation monitoring tool. *NASA Tech Briefs, 32*, 63-64

McKellip, R.D., K.W. Ross, J.P. Spruce, J.C. Smoot, R.E. Ryan, G.E. Gasser, D.L. Prados, and R.D. Vaughan (2010). Phenological Parameters Estimation Tool. *NASA Tech Briefs, 34,* 49-50

McMaster, G.S. (2005). Phytomers, phyllochrons, phenology and temperate cereal development. *Journal of Agricultural Science, 143*, 137-150

Mithal, V., A. Garg., S. Boriah, M. Steinbach, V. Kumar, C. Potter, S. Klooster, and J.C. Castilla-Rubio (2011). Monitoring global forest cover using data mining. *ACM Transactions on Intelligent Systems and Technology, 2,* 36:1-36:24

Miura, T., A.R. Huete, W.J.D. van Leeuwen, and K. Didan (1998). Vegetation detection through smoke-filled AVIRIS images: An assessment using MODIS band passes. *Journal of Geophysical Research, 103*, 32.001-32.011

Morisette, J., A. Richardson, A. Knapp, J. Fisher, E. Graham, J. Abatzoglou, B. Wilson, D. Breshears, G. Henebry, J. Hanes, and L. Liang (2009). Tracking the rythm of the seasons in the face of global change: phenological research in the 21st Century. *Frontiers in Ecology and the Environment, 7,* 253-260

Moulin, S., L. Kergoat, N. Viovy, and G. Dedieu (1997). Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements. *Journal of Climate, 10*, 1154-1170

Muchoney, D.M., J. Borak, and A. Strahler (1996). Global land cover classification validation issues and requirements. *Proceedings of the 1996 IEEE International Geoscience and Remote Sensing Symposium, 1*, 233-235

NASA History Program Office (2014). A Brief History of NASA. *Available at* http://history.nasa.gov/factsheet.htm (accessed in February 24, 2014)

NASA Missions (2014). International Space Station. *Available at* http://www.nasa.gov/mission_pages/station/cooperation/ (accessed in February 24, 2014)

NASA Science Missions (2014). TIROS – Television Infrared Observation Satellite Program. *Available at* http://science1.nasa.gov/missions/tiros/ (accessed in February 24, 2014)

NOAA (2014). National Oceanic and Atmospheric Administration: Environmental Visualization Laboratory. *Available at* http://www.nnvl.noaa.gov/green.php (accessed in March13, 2014)

NRC, National Research Council (1997). Precision Agriculture in the 21st Century. *National Academy Press*, Washington D.C., 149

Pettorelli, N., J. Vik, A. Mysterud, J. Gaillard, C. Ticker, and N. Stenseth (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution, 20,* 503-510

Pôças, I., M. Cunha, A.R.S. Marçal, and L.S. Pereira (2011). An Evaluation of changes in a mountainous rural landscape of Northeast Portugal using remotely sensed data. *Landscape and Urban Planning, 101*, 253–261

Potter C., P.N. Tan, M. Steinbach, S. Klooster, V. Kumar, R. Myneni, and V. Genovese (2003). Major disturbance events in terrestrial ecosystems detected using global satellite data sets. *Global Change Biology, 9,* 1005-1021

Prados, D., R.E. Ryan, and K.W. Ross (2006). Remote Sensing Time Series Product Tool. *AGU Fall Metting 2006*.

Reddy, B.V.S., P.S. Reddy, and F. Bidinger (2003). Crop management factors influencing yield and quality of crop residues. *Field Crops Research, 84*, 57-77

Reed, B.C., J.F. Brown, D. VanderZee, T.R. Loveland, J.W. Merchant, and D.O. Ohlen (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science, 5,* 703-714

Reichhardt, T. (2006). The first photo from space. *AIR&SPACE Magazine. Available at http://www.airspacemag.com/space/the-first-photo-from-space-13721411/* (accessed in June 8, 2014)

Richardson, A.J., and C.L. Wiegand (1977). Distinguishing vegetation from soil backgroud information. *Photogrammetric Engineering and Remote Sensing, 43*, 1541-1552

Roerink, G., M. Menenti, and W. Verhoef (2000). Reconstructing cloudfree NDVI composites using Fourier analysis of time-series. *International Journal of Remote Sensing, 21*, 1911-1917

Rouse, J.W., R.H. Haas, J.A. Schell, and D.W. Deering (1974). Monitoring vegetation systems in Great Plains with ERST. *Proccedings of the 3rd ERST Symposium, NASA SP-351*, 309-317

Roy, D. (2000). The impact of misregistration upon composited wide field of view satellite data and implications for change detection. *IEEE Transactions on Geoscience and Remote Sensing, 38,* 2017-2032

Running, S., C. Justice, V. Salomonson, D. Hall, J. Barker, Y. Kaufmann, A. Strahler, A. Huete, J.P. Muller, and V. Vanderbilt (1994). Terrestrial remote sensing science and algorithms planned for EOS/MODIS. *International Journal of Remote Sensing, 15*, 3587-3620

Running, S.W., R.R. Nemani, F.A. Heinsch, M. Zhao, M. Reeves, and H. Hashimoto (2004). A continuous satellite-derived measure of global terrestrial primary production. *BioScience, 54*, 547-560

Sakamoto, T., M. Yokozawa, H. Toritani, M. Shibayama, N. Ishitsuka, and H. Ohno (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment, 96,* 366-374

Schellberg, J., M.J. Hill, R. Gerhards, M. Rothmund, and M. Braun (2008). Precision Agriculture of grassland: Applications, perspectives and constraints. *European Journal of Agronomy, 29*, 59-71

Schulze, E.D., D. Mollicone, F. Achard, G. Matteucci, S. Federici, H.D. Eva, and R. Valentini (2003). Making deforestation pay under the Kyoto Protocol? *Science, 299,* 1669

Schwartz, M.D., and T.R. Karl (1990). Spring phenology: nature's experimet to detect the effect of green-up on surface maximum temperatures. *Monthly Weather Review, 118*, 883-890

Sellers, P.J. (1985). Canopy reflectance, photosynthesis, and transpiration. *International Journal of Remote Sensing, 6,* 1335-1372

SIC (Satellite Imaging Corporation) (2014). Characterization of Satellite Remote Sensing Systems. *Available at http://www.satimagingcorp.com/characterization-of-satellite-remote-sensing-systems.html* (accessed in February 24, 2014)

Sripada, R.P., R.W. Heinger, J.G. White, and A.D. Meijer (2006). Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agronomy Journal, 98, 9*68-977

Strahler, A.H., L. Boschetti, G.M. Foody, M.A. Friedl, M.C. Hansen, M. Herold, P. Mayaux, J.T. Morisette, S.V. Stehman, and C.E. Woodcock (2006). Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. *Global Observation of Forest and Land Cover Dynamics (GOFC-GOLD) Report no. 25, EUR 22156- EN.*

Sweets, D., B. Reed, J. Rowland, and S. Marko (1999). A weighted least-squares approach to temporal smoothing of NDVI. *Proceedings of the American Society for Photogrammetry and Remote Sensing Conference,* 526-536

Tan, B., J.T. Morisette, R.E. Wolfe, F. Gao, G.A. Ederer, J. Nightingale, and J.A. Pedelty (2011). An Enhanced TIMESAT algorithm for estimating vegetation phenology metrics from MODIS data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4*, 361-371

Teillet, P.M., K. Staenz, and D.J. Williams (1997). Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions. *Remote Sensing of Environment, 61,* 139-149

Tucker, C.J., J.R.G. Townshend, and T.E. Goff (1985). African land-cover classification using satellite data. *Science, 227,* 369-375

Udelhoven, T. (2011). TimeStats: A software tool for the retrieval of temporal patterns from global satellite archives. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4*, 310-317

van Leeuwen, W., B. Orr, S. Marsh, and S. Herrmann (2006). Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications. *Remote Sensing of Environment, 100*, 67-81

Velleman, P. (1980). Definition and comparison of robust nonlinear data smoothing algorithms. *Journal of the American Statistical Association, 75*, 609-615

Venkataratnam, L. (2001). Remote senisng and GIS in agricultural resources management. *Proceedings of the 1st National Conference on Agro-Informatics*, 20-29

Verbesselt, J., P. Jonsson, S. Lhermitte, I. Jonckheere, J. van Aardt, and P. Coppin (2012). Relating time-series of meteorological and remote sensing indices to monitor vegetation moisture dynamics. *C.H. Chen (Ed.), Signal and Image Processing for Remote Sensing.* CRC Press, 129-146.

Vermote, E.F., and A.Vermeulen (1999). Atmospheric correction algorithm: Spectral reflectances (MOD09). *MODIS Algorithm Technical Background Document, NASA contract NAS5-96062*

Vermote, E.F., N.Z. El-Saleous, and C.O. Justice (2002). Atmospheric correction of MODIS data in the visible to middle infrared: first results. *Remote Sensing of Environment, 83*, 97-111

Viovy, N., O. Arino, and A. Belward (1992). The Best Index Slope Extraction (BISE): a method for reducing noise in NDVI time-series. *International Journal of Remote Sensing, 13,* 1585-1590

Wardlow, B.D., S.L. Egbert, and J.H. Kastens (2007). Analysis of time-series MODIS 250m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment, 108*, 290-310

Watson, D. (1947). Comparative physiological studies in the growth of field crops. I. variation in net assimilation rate and leaf area between species and varieties, and within and between years. *Annals of Botany 11*, 41-76.

Weier, J., and D. Herring (2000). Measuring Vegetation (NDVI & EVI). *Available at http://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_1.php* (accessed in February 28, 2014)

Woolley, J.T. (1971). Reflectance and transmittance of light by leaves. *Plant Phisiology, 47*, 656-662

Xiao, X., B. Braswell, Q. Zhang, S. Boles, S. Frolking, and B. Moore III (2003). Sensitivity of vegetation indices to atmospheric aerosols: continental-scale observations in Northern Asia. *Remote Sensing of Environment, 84,* 385-392

Xiao, X., D. Hollinger, J. Aber, M. Goltz, E.A. Davidson, Q. Zhang, and B. Moore III (2004a). Satellite-based modeling of gross primary production in an evergreen needleaf forest. *Remote Sensing of Environment, 89*, 519-534

Xiao, X.M., Q. Zhang, B. Braswell, S. Urbanski, S. Boles, S. Wofsy, B. Moore III, and D. Ojima (2004b). Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment, 91,* 256-270

Yoshioka, H., T. Miura, H. Yamamoto, and A. Huete (2002). A technique of inter-sensor VI translations using EO-1 Hyperion data to minimize systematic diferences in spectral band-pass filters. *Proceedings of the IEEE 2002 International Geoscience and Remote Sensing Symposium, 4,* 2211-2213

Zeng, H., G. Jia, and B.C. Forbes (2013). Shifts in Artic phenology in response to climate and anthropogenic factors as detected from multiple satellite time series. *Environmental Research Letters, 8,* 1-12.

Zhang, X., M. Friedl, C. Schaaf, A. Strahler, J. Hodges, F. Gao, B. Reed, and A. Huete (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment, 84*, 471-475

Zhang, X., M. Friedl, C. Schaaf, and A. Strahler (2004). Climate controls on vegetation phenological parameters in northern mid- and high latitudes inferred from MODIS data. *Global Change Biology, 10,* 1133-1145.

Zhou, L., R.K. Kaufmann, Y. Tian, R.B. Myneni, and C.J. Tucker (2003). Relation between interannual variations in satellite measures of northern forest greenness and climate between 1982 and 1999. *Journal of Geophysical Research, 108, ACL3-1-ACL3-16.*

APPENDICES

APPENDIX 1

PhenoSat - A tool for vegetation temporal analysis from satellite image data

PHENOSAT – A TOOL FOR VEGETATION TEMPORAL ANALYSIS FROM SATELLITE IMAGE DATA

Arlete Rodrigues, André R.S. Marcal, Mário Cunha

Centro de Investigação em Ciências Geo-Espaciais, Faculdade de Ciências, Universidade do Porto

ABSTRACT

The availability of temporal satellite image data has increased considerably in recent years. A number of satellite sensors currently observe the Earth with high temporal frequency thus providing a tool for monitoring/understanding the Earth-surface variability more precisely, for several applications such as the analysis of vegetation dynamics. However, the extraction of vegetation phenology information from Earth Observation Satellite (EOS) data is not easy, requiring efficient processing algorithms to properly handle the large amounts of data gathered.

The purpose of this work is to present a new, easy-to-use software tool that produces phenology information from EOS vegetation temporal data – PhenoSat. This paper describes PhenoSat, focusing on two new features: the determination of the beginning and maximum of a double growth season, and the selection of a temporal sub-region of interest in order to reduce and control the data evaluated.

Index Terms— Remote Sensing; Vegetation Index; Time-series; Phenology; Vegetation Dynamics.

1. INTRODUCTION

The number of satellite sensors with capability to acquire temporal image data has considerably increased in the last few years. The most important of these sensors are NOAA AVHRR, SPOT VEGETATION, ENVISAT MERIS, Metop-AVHRR and MODIS. With the increased number of sensors and improved data availability, it is now considerably easier to collect high temporal resolution image datasets, allowing for the monitorization of the Earth-surface to be done more frequently and with increased precision.

Phenological studies are very important to understand ecosystem functioning and associated seasonal patterns of carbon, water and energy fluxes, as well as to quantify crop response to climate variability and agronomic practices. The easier access to vegetation temporal data provides a continuous high frequency range of observations, and thus a more detailed crop and vegetation monitoring is now possible. However, the adequate processing of the large data volumes obtained is a difficult and time consuming process. Therefore, the development of software tools that allow for an efficient, semi-automatic processing is highly necessary.

There are a number of software tools available to process vegetation temporal data, such as TiSeG [1], TSPT [2], TIMESAT [3] and PPET [4]. These tools enable the extraction of phenology information, but in general do not permit the extraction of phenology for a double growth season, neither a selection of the data interval to process.

The most common approach to extract vegetation parameters (phenological phases) uses deterministic models based on regression analysis. In [5] it is shown that the best methods were developed by [6] and by [7]. Although most of these models provide satisfactory parameters estimation, they present limitations when applied to different years, areas and vegetation types.

The main objective of this work is to present the software tool PhenoSat for the extraction of phenology information from vegetation indices temporal data, addressing the flaws encountered in the existing software packages.

This paper presents the methodologies used for processing temporal series of vegetation index imagery and the main features includes in the software tool developed.

1.1. Data

The Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) are remote sensed vegetation indices (VI) commonly used in vegetation dynamics studies. NDVI is calculated from the reflectance in RED and near-infrared (NIR) spectral bands [8] using Eq.1.

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (1)$$

$$EVI = G.\frac{(NIR - RED)}{(L + NIR + C_1 RED + C_2 BLUE)} \quad (2)$$

The EVI improves upon the quality of the NDVI product, correcting some distortions in the reflected light caused by particles in the air as well as the ground cover below the vegetation. When viewing rainforests and other areas of the Earth with large amounts of chlorophyll, the EVI has the advantage that it does not become saturated as easily as the NDVI. EVI is calculated using Eq.2, where *NIR*, *RED* and *BLUE* are the surface reflectance in near-infrared, red and blue spectral bands, G is a gain factor, L is a canopy background adjustment term, and C_1 and C_2 are the coefficients of the aerosol resistance term, which use the blue band to correct for aerosol influences in the red band [9].

For agricultural and vegetation condition monitoring, clouds are partially screened from NDVI and EVI images by producing Maximum Value Composites (MVC) over a specific period. For this period, the NDVI (EVI) values are analyzed, on a pixel-by-

 978-1-4577-1203-6/11/$26.00 ©2011 IEEE

pixel basis, and the highest value is retained for each pixel location. When all pixels have been evaluated, a MVC image is obtained. In [10] it is shown that the MVC imagery is highly related to green vegetation dynamic, and the common problems present in single-date remote sensing studies, as cloud contamination, atmospheric attenuation and observation geometry are minimized using MVC.

VI satellite data are easily available and can be searched and ordered with reduced or no cost over the internet. These datasets are organized in two-dimensional images, with the time acting as a 3rd dimension in the data cube (Fig.1). For an observation *t*, for pixel location *(j,k)*, $N_t(j,k)$ gives the VI value at time *t* [11]. The set of observations $N_{ti}(j,k)$, with i=1,…,N, is the time-series for pixel *(j, k)*.

Time-series of VI can be used to obtain information on seasonal vegetation development. This information is essential to understand the characteristics and behavior of vegetation dynamics, thus allowing a more precise knowledge and monitorization [12] [13].

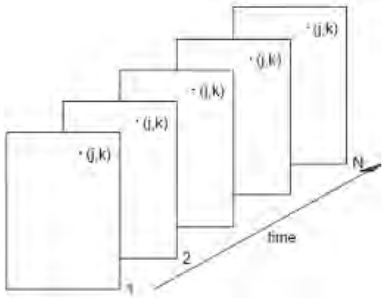


Fig.1. Vegetation indices satellite data organized in two-dimensional images.

2. METHODOLOGY

The software presented here – PhenoSat – was developed using Matlab [14]. It receives as input a VI temporal data cube (a set of VI images). It is important to understand that VI values are not only dependent of vegetation density and type, they are also influenced by observation geometry, atmospheric and illumination conditions, although the MVC technique used minimizes the influence of these undesired factors.

The final temporal VI data has nevertheless some noise due to the contribution of some parameters that cannot be fully eliminated by the MVC process. To reduce this noise and its influence in the results, a moving median filter [15] is after applied to remove rare events such as rapid shocks or other anomalies. Usually the application of the median filter is not sufficient to properly remove the noise present in temporal VI data. To remove occasional outliers, a Savitzky-Golay (SG) filter [16] can be applied. The SG filter coefficients are derived by performing an unweighted linear least square fit using a polynomial of a given degree. In general, higher degree polynomials can more accurately capture the heights and widths of narrow peaks, but perform poorly at smoothing wider peaks.

Field-based ecological studies have proved that vegetation phenology tends to follow relatively well-defined temporal patterns. The beginning of growth is followed by a rapid growth, stabilizing in the maximum leaf area. The transition between senescence follows a similar trend, but in a reverse pattern. Thus, it is possible to represent the vegetation growth by a logistic function [17] [18], such as the period between senescence and dormancy.

Temporal variations of satellite NDVI data can be modeled by a double logistic function of the type presented in Eq.3, where *t* is the Julian day, VI_t is the VI value at time *t*, *k* is related to high asymptotical value of *VI*, *c* and *d* denote the slopes at the "left" and "right" inflection points, respectively, and *p* and *e* are the dates of these inflection points. VI_w and VI_{w1} are the VI values before the budbreak and during the dormancy after the leaf fall, respectively. This double logistic function is similar to others already developed ([17] [6]) but the *k* parameter insures the continuity between the vegetation growth and senescence parts, even when they differ in shape [19] [20].

$$VI_t = VI_w + \frac{k}{1+\exp[-c(t-p)]} - \frac{k+VI_w-VI_{w1}}{1+\exp[-d(t-e)]} \qquad (3)$$

As an illustration, Fig.2 shows the NDVI values optimized after a double logistic function application and a schematic procedure for the calculation of the transition dates. As shown in Fig.2, it is possible to identify the principal phenological phases (Green-up (Gu), Maturity (M), Senescence (Se) and Dormancy (Dm)) from the rate of curvature change of Eq.3, using its minima and maxima. To identify the inflexion points (IP_L and IP_R) the global maximum and minimum of the curvature are used.

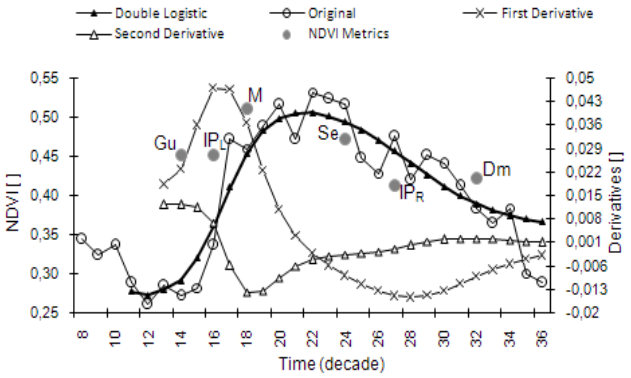


Fig.2. Schematic representation of NDVI Metrics calculated based on maxima and minima rate of curvature change. Global maximum and minimum of the curvature estimate the inflexion points.

$$\Delta = VI_{t+1} - VI_t \qquad (4)$$

$$VI_{t+1} < VI_t \;\&\; VI_{t+2} < VI_{t+1} \qquad (5)$$

The PhenoSat program permits to deal with double-crop growth season and/or vegetation re-growths during the seasons. To estimate the time occurrences of this re-growth, for a specific year *N*, the data after dormancy in year *N* and all data in year *N+1* are considered. Eq.4 and Eq.5, where *t* is the time occurrence and VI_t is the VI value at time *t*, allow for the calculation of the beginning

978-1-4577-1203-6/11/$26.00 ©2011 IEEE

of re-growth and its maximum value. The beginning occurs when $\Delta > 0$ for three or more distinct points and the maximum VI value occurs at a time *t* that verifies Eq.5.

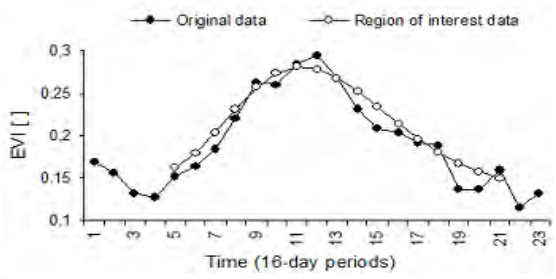


Fig.3. Example of an EVI temporal series for a year: original data and adjusted to a subset of the data.

Two more options are available in PhenoSat: the execution of an upper envelope [6] to enhance the spring and summer periods, and the selection of a region of interest allowing a reduction of the temporal data volume to be processed.

The region of interest can be selected manually by the user, or automatically by PhenoSat. The automatic selection of the temporal region of interest is done by PhenoSat, by first finding the maximum value of the data and searching for the point where a significant increase (or abrupt decrease) is verified at the left of that maximum. That point corresponds to the initial position of the region of interest. Afterwards, to determine the final position the program proceeds in a similar way, but evaluates the data at the right of the maximum (Fig.3). The data between initial and final positions are used to compute the phenological phases, with the remaining data being discarded.

The outputs of PhenoSat are two text (on MSExcel) files containing information about data processing (original and fitting data) and the vegetation dynamic information (phenology phases such as green-up, maturity, senescence, dormancy, and beginning and maximum of a double growth season if it occurs).

3. TEST RUN

PhenoSat was tested in a vineyard and semi-natural mountain meadows both in the North of Portugal. AQUA MODIS EVI satellite images, with 16-days temporal (23 images per year) and 250 meters spatial resolutions, were acquired from Land Processes DAAC User Services, USGS Earth Resources Observation and Science (EROS) Center. All available data for this site were used – 6 years (2003-2008).

A sub-image representing a specific test site in Régua (latitude 41º.167, longitude -7º.783) region was selected. The EVI values were extracted for all years for this sub-image, thus obtaining the EVI time-series.

As the EVI values are not distributed in a [0-1] range, the first step was to scale those values using the information provided by MODIS [21] image information. The VI data was exported to a text file table with 6 columns (years) and 23 rows (16 day MVCs). This text file was selected as input file in the PhenoSat interface. Median and SG filters were applied as an upper envelope which enhances the rise and fall of the curve.

Table1. Time occurrences (in 16-day periods) of the phenological phases obtained from EVI time-series in Régua vineyard region, between 2003 and 2008.

	2003	**2004**	**2005**	**2006**	**2007**	**2008**
Green-up	5.72	6.81	5.98	6.02	7.15	7.18
Inflexion1	7.63	8.14	7.26	7.94	8.32	8.68
Maturity	10.04	9.64	8.64	11.32	9.54	10.41
Senescence	11.54	11.73	11.15	12.84	14.46	12.21
Inflexion2	15.34	15.29	15.01	14.69	16.41	15.72
Dormancy	18.45	18.27	17.86	16.17	18.30	19.25
Maximum	11.2	11.04	10.46	11.73	11.63	11.68

To minimize the processing time and focus on the relevant data, a region of interest was defined with initial and final positions as 5 (March) and 21 (November). The 7 parameters of the double logistic function were estimated using the Levenberg-Marquardt method [22], and the result was used to fit the data in the interval (5-21). Finally, the fitted data obtained were used to compute the phenological phases (green-up, inflexion1, maturity, senescence, inflexion2 and dormancy).

The process described here was applied to all available years (2003-2008). Table1 presents the time occurrences of the phenological phases obtained from PhenoSat for all 6 years. The mean value of R-square, for all years, was estimated in 0.99 which proves the efficiency of the method to fit the EVI time-series used.

As the vineyard crop does not verify a double growth season in the same year, a semi-natural mountain meadows (*"lameiros"*) crop, in Northwest Portugal was selected for testing. In some years this type of crop verifies a re-growth, so it is ideal to test the ability of the software in its detection. If the re-growth does not occurs in the same year, the value of the beginning and maximum refers to the following year, and the maximum of the re-growth in a year can be coincident to the maximum vegetation growth in the following year. Fig.4 shows the NDVI data acquired in 2006, 2007 and 2008 in a *"lameiros"* crop, where it is possible to see the values of re-growth obtained in two years (2006 and 2007).

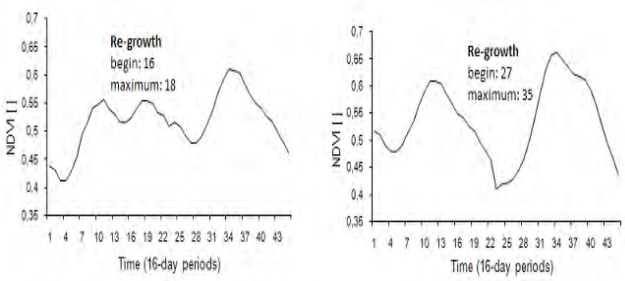


Fig.4. NDVI data after SG filter application, from "lameiros" crop in 2006 and 2007 (left) and in 2007 and 2008 (right), and the re-growth in 2006 and 2007, respectively.

To evaluate the Phenosat software and its results, different tests using various test sites, various VI time-series and various sensors were done, and a comparison was made between the phenology information obtained in the field and those computed by the software. The results in the field and obtained by PhenoSat were found to be very similar for vineyard [20] and for lameiros [23].

5. CONCLUSIONS

This work presents a new software tool (PhenoSat), developed in Matlab, which allows for information about phenology phases to be obtained from vegetation index (VI) temporal data, using an easy-to-use semi-automatic process. This software tool focuses in two new and relevant features: the ability to obtain information about a double growth season and the option available to select a temporal sub-region of interest.

PhenoSat was tested for different VI time-series acquired from different sensors, for different years and in different geographical areas. A test case using MODIS AQUA data was presented here. The results obtained from PhenoSat proved to be very similar to measurements made in field. Thus, satellite VI data processed by PhenoSat has high capability to predict phenology phases, and has the advantage of doing it in a short time and in a relatively simple way.

The PhenoSat software tool is available for public use freely at http://www.fc.up.pt/LamSat_XXI.

6. ACKNOWLEDGMENTS

Arlete Rodrigues has a PhD grant (SFRH / BD / 62189 / 2009) by FCT (Fundação para a Ciência e a Tecnologia, Portugal). This work is also part of the FCT project LamSatXXI, FCOMP-01-0124-FEDER-006996.

7. REFERENCES

[1] R.R. Colditz, C. Conrad, T. Whermann, M. Schmidt and S. Dech, "TiSeG: A flexible software tool for time-series generation of MODIS data utilizing the quality assessment science data set", *IEEE Transactions on Geoscience and Remote Sensing 46 (10),* pp. 3296-3308, 2008.

[2] R. McKellip, D. Prados, R. Ryan, J. Spruce, G. Gasser and R. Greer, "Remote sensing time-series analysis, a vegetation monitoring tool ", *NASA Tech Briefs,* 2008.

[3] P. Jonsson and L. Eklundh, "TIMESAT – a program for analyzing time-series of satellite sensor data", *Computers&Geosciences 30 (8),* pp. 833-845, 2004.

[4] K.W. Ross, G. Gasser and B. Spiering, "Feasibility of estimating relative nutrient contributions of agriculture using Modis time-series", *NASA Technical Reports Server, Stennis Space Center, Mississipi,* 2008.

[5] J.N. Hird and G. McDermid, "Noise reduction of NDVI time-series: An empirical comparison of selected techniques", *Remote Sensing of Environment 113 (1),* pp.248-258, *2009*.

[6] P.S.A. Beck, C. Atzberger, K.A. Hogda, B. Johansen and A.K. Skidmore, "Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI", *Remote Sensing of Environment 100 (3),* pp.321-334, *2006*.

[7] P. Jonsson and L. Eklundh, "Seasonality extraction by function-fitting to time-series of satellite sensor data", *IEEE Transactions on Geoscience and Remote Sensing 40 (8),* pp.1824-1832, *2002*.

[8] A. Huete, C. Justice and W.V. Leeuwen, "MODIS vegetation index (MOD13): algorithm theoretical basis document", 1999.

[9] Z. Jiang, A.R. Huete, K. Didan and T. Miura, "Development of a two-band enhanced vegetation index without a blue band", *Remote Sensing of Environment 112 (10),* pp.3833-3845, 2008.

[10] C.J. Tucker, J.U. Hielkema and J. Roffey, "The potential of satellite remote sensing of ecological conditions for survey and forecasting desert-locust activity", *International Journal of Remote Sensing of Environment 6 (1),* pp.127-138, 1985.

[11] J.W.J Rouse, R.H. Haas, J.A. Schell and D.W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS", *Third Symposium, NASA SP-351, Washington DC,* pp.309-317, 1974.

[12] C.J. Tucker, C.O. Justice and S.D. Prince, "Monitoring the grasslands of the Sahel 1984-1985", *International Journal of Remote Sensing 7 (11),* pp.1571-1581, 1986.

[13] J.P. Malingreau, "Global vegetation dynamics: satellite observations over Asia", *International Journal of Remote Sensing 7 (9),* pp.1121-1146, 1986.

[14] D.J. Higham, N.J. Higham, *Matlab Guide,* SIAM, Philadelphia, 2000.

[15] P. Gonçalves, H. Carrão and M. Caetano, "Parametric model for intra-annual reflectance time seires", *Proccedings of the 26th EARSeL Symposium, New Developments and Challenges in Remote Sensing, edited by Z. Bochenek, Millpress,* pp.367-375, 2007.

[16] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing, third edition,* pp.766-772, 2007.

[17] X. Zhang, M.A. Friedl, C.B. Schaaf, A.H. Strahler, J.G.F. Hodges, F. Gao, B.C. Reed and A.R. Huete, "Monitoring vegetation phenology using MODIS", *Remote Sensing of Environment 84 (3),* pp.471-475, 2003.

[18] D. Villegas, N. Aparicio, R. Blanco and C. RoYo, "Biomass accumulation and main stern elongation of durum wheat grown under Mediterranean conditions", A*nnuals of Botany 88 (4),* pp.617-627, *2001*.

[19] A. Fischer, "A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters", Remote *Sensing of Environment 48 (2),* pp.220-230, *1994*.

[20] M. Cunha, A.R.S. Marçal and A. Rodrigues, "A Comparative Study of Satellite and Ground-Based Vineyard Phenology", *Proceedings of the 29th EARSeL Symposium, Imagin [e,g] Europe,* pp.68-77, *2010*.

[21] MODIS Website, https://lpdaac.usgs.gov/lpdaac/products/ (available at 27May2011).

[22] D.W. Marquardt, "An algorithm for least-squares estimation of non-linear parameters", *Journal of the Society for Industrial and Applied Mathematics 11 (2)*, pp.431-441, 1963.

[23] M. Cunha, I. Poças, A.R.S. Marçal, A. Rodrigues and L. Pereira, "Evaluating MODIS vegetation indices using ground based measurements in mountain semi-natural meadows of Northeast Portugal", *Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS),* pp.1525-1528, *2010*.

APPENDIX 2

Phenology parameter extraction from time-series of satellite vegetation index data using PhenoSat

PHENOLOGY PARAMETER EXTRACTION FROM TIME-SERIES OF SATELLITE VEGETATION INDEX DATA USING PHENOSAT

Arlete Rodrigues, Andre R.S. Marcal, Mario Cunha

Centro de Investigação em Ciências Geo-Espaciais
Faculdade de Ciências, Universidade do Porto, Portugal

ABSTRACT

PhenoSat is an experimental software tool that extracts phenological information from satellite vegetation index time-series. Temporal satellite NDVI data provided by VEGETATION sensor from three different vegetation types (Vineyard, Closed Deciduous Forest and Deciduous Shrubland with Sparse Trees) and for different geographical locations were used to test the ability of the software in extracting vegetation dynamics information. Six noise reduction filters were tested: piecewise-logistic, Savitzky-Golay, cubic smoothing splines, Gaussian models, Fourier series and polynomial curve fitting. The results showed that PhenoSat is an useful tool to extract phenological NDVI metrics, providing similar results to those obtained from field measurements. The best results presented correlations of 0.89 ($n=6$; $p<0.01$) and 0.71 ($n=6$; $p<0.06$) for the green-up and maximum stages, respectively. In the fitting process, the polynomial and Gaussian algorithms over smoothed the peak related with a double-growth season, the opposite to the other methods that could detect more accurately this peak.

Index Terms— PhenoSat, Phenology, NDVI, SPOT VGT, Time-series

1. INTRODUCTION

Earth Observation Satellites provide a cost-effective mean for assessing the vegetation dynamics in agricultural and natural ecosystems. In recent years the number of satellite sensors with capability to acquire temporal data has increased considerably. Most of these sensors provide temporal vegetation data associated to a specific vegetation index (VI). Time-series satellite imagery can provide a synoptic view of vegetation dynamics by measuring surface reflectance at regular time-intervals that may be used to extract the phenological information required to better analyze and understand global changes in vegetation. The evolution of the VI through time establishes temporal profiles that can be analyzed to extract meaningful parameters about the vegetation dynamics. Although VI time-series data from different sensors are widely available, there is still a considerable gap between the raw data and the information actually needed.

PhenoSat is an automatic tool to analyze time-series data and extract phenology information in an easy and fast manner. A preliminary application of PhenoSat was used on vegetation dynamics studies and is reported in [1]. Nevertheless, much work has to be done to accurately calibrate the use of the PhenoSat for different vegetation cover types and geographic locations.

This paper presents the results obtained from a number of algorithms used to extract phenological information from satellite VI time-series. These algorithms were implemented in PhenoSat and, using NDVI data from SPOT VEGETATION (VGT), we illustrate the performance of PhenoSat in 3 different vegetation types.

2. MATERIALS AND METHODS

The algorithms implemented in PhenoSat allow to extract a number of phenological information parameters, such as: green-up, maturity, senescence, dormancy and a double-growth (regrowth) occurrence, if it happens.

2.1. Dataset and study area

Ten-days NDVI composite times-series obtained by SPOT VEGETATION were used for three test sites located in Africa (2 sites) and Portugal (1 site). Three different vegetation types were analysed: vineyard (VIN; Portugal), Closed Deciduous Forest (CDF; Africa) and Deciduous Shrubland with Sparse Trees (DSST; Africa). These land cover types were selected for the available ground-based information as well as for their different growth patterns.

The GLC2000 map for Africa was used to select the test sites for CDF and DSST. The PhenoSat derived phenology parameters obtained for each algorithm were compared with ground-based phenological observations for vineyard over the years 2000-2005. Table I presents the data description used in the different experiments carried out.

Table I. Satellite data description used in this paper experiments.

Land Cover Type	Code	Location	Satellite product	Time-series period	Size (pixels)
Vineyard (*)	VIN	Portugal	SPOT VGT (1km; 10 days)	2000 to 2005	9
Closed Deciduous Forest	CDF	Africa	SPOT VGT (1km; 10 days)	2000	2408
Deciduous Shrubland with Sparse Trees	DSST	Africa	SPOT VGT (1km; 10 days)	2000	1740

*Ground-based measurements available.

2.2. Methodology

PhenoSat receives as input the NDVI temporal profiles extracted for a given test site. In order to analyze the temporal profiles and compute the NDVI metrics related with the phenological vegetation, one of six algorithms can be used to fit the data: Savitzky-Golay filter (SG), piecewise-logistic (DL), cubic smoothing splines (CSS), polynomial curve fitting (PCF), Fourier series (FS) or Gaussian model (GM). The SG algorithm performs a local polynomial regression on a series of values to determine the smoothed value for each point [2].

Some studies showed that vegetation dynamics tends to follow a well-defined growth temporal pattern and the vegetation cycle can be represented by a piecewise-logistic algorithm (DL) [3, 4]. The DL used in this work is based on 7 parameters to fit the vegetation data, as described in [1, 4].

The CSS algorithm fits a spline or smooth piecewise-polynomial algorithm, and a smoothing parameter determines just how closely the smoothing spline adheres to the given data [6].

The PCF [7] applies a polynomial of a given degree to fit the data. The higher the degree, the closer the fitting curve will be to the given data, although this should be done only up a certain degree.

The Fourier series [8] is a sum of sine and cosine functions that describes a periodic signal. It is represented in either the trigonometric form or the exponential form (Eq.1), where a_0 models a constant (intercept) term in the data and is associated with the $i = 0$ cosine term, w is the fundamental frequency of the signal, n is the number of terms (harmonics) in the series, and $1 \leq n \leq 8$.

The Gaussian model [9] fits peaks and is given by Eq.2, where a is the amplitude, b the centroid (location), c is related to the peak width and n is the number of peaks to fit ($1 \leq n \leq 8$).

$$y = a_0 + \sum_{i=1}^{n} a_i \cos(nwx) + b_i sen(nwx) \quad (1)$$

$$y = \sum_{i=1}^{n} a_i e^{\left[-\left(\frac{x-b_i}{c_i}\right)^2\right]} \quad (2)$$

2.3. Evaluation strategy

A SG filter with 1st degree polynomial and frame size 5 was applied. These parameters values were chosen because they smoothed the data just enough to remove the outliers and spikes, without suppressing natural variations of the vegetation. The results obtained from the SG filter were then fitted by the CSS, GM, PCF, FS or DL algorithms. The performance of CSS, FS, GM and PCF algorithms depends on a smoothing parameter, which affects the adherence of the fitting curve to the original data. PhenoSat was used to test the sensitivity of each algorithm by varying the smoothing parameter through a range of possibilities.

Seven satellite based phenological stages (NDVI metrics) were determined: green-up (Gup), "left" and "right" inflection points (LIP and RIP), maturity (Mat), maximum (Max), senescence (Se) and dormancy (Dm). The Max stage was defined as the maximum NDVI value observed in the data interval delimited by LIP and RIP. The other six phenological stages were determined based on the maxima and minima of the first and second derivatives of the fitted data. As an illustration, Fig. 1 shows the NDVI values optimized after DL application and a schematic procedure for the calculation of the NDVI metrics.

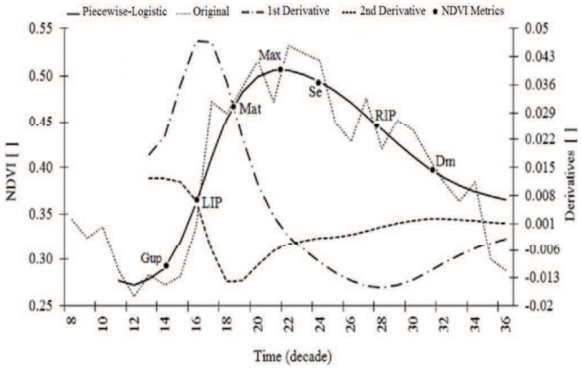


Fig. 1. Representation of NDVI Metrics calculated based on maxima and minima rate of curvature change. Global maximum and minimum of the curvature estimate the inflexion points.

3. RESULTS

The NDVI profiles of the year 2000 for the three test sites selected are presented in Fig. 2. These profiles were determined using the median NDVI computed for each test

site and for each image available in the year 2000 (36 images).

The CDF profile verifies a double-growth occurrence in the same year, and the DSST profile presents a marked difference between summer and winter growth without in-season regrows. The grapevine in Portugal can have a not so well-defined profile as this culture has a long dormancy period and a discontinuous canopy with understory vegetation growth.

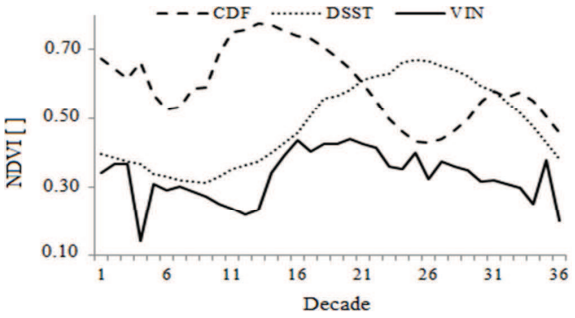


Fig. 2. Temporal NDVI profiles for Closed Deciduous Forest (CDF), Deciduous Shrubland with Sparse Trees (DSST) and Vineyard (VIN), for the year 2000.

3.1. PhenoSat derived NDVI metrics

The original and the curves fitted by PhenoSat for CDF vegetation type, using the GM, PCF and CSS algorithms, are presented in Fig. 3. The DL, FS and SG fitting results are similar to those obtained using the CSS algorithm. Fig. 3 shows PCF and GM with considerable differences in the fitted curves, comparatively to the CSS method. The biggest difference was on the regrowth period where they were unable to fit the second peak of the original data, over smoothing this occurrence. The CSS could more accurately detect the peaks and transitions on the original data (with an r-square of 0.93), which was also verified in the DL (0.98), SG (0.93) and FS (0.91) algorithms.

Although there are differences in the fitting process, the phenological values obtained for the regrowth are identical for all algorithms. The beginning and maximum of this period occurred on decades 27 and 32, respectively. For the main growing season, the phenological results were very similar for DL, SG and CSS, being the PCF the most distinct with an early green-up occurrence (Table II).

For the DSST vegetation type, the phenological stages obtained from PhenoSat presented a greater similarity over the six algorithms. However, the PCF and FS showed a slight difference at the beginning of the growth season, having a green-up occurrence earlier than on the other methods (Table III). The SG and CSS presented identical results and GM was the algorithm with the lower r-square value (0.96).

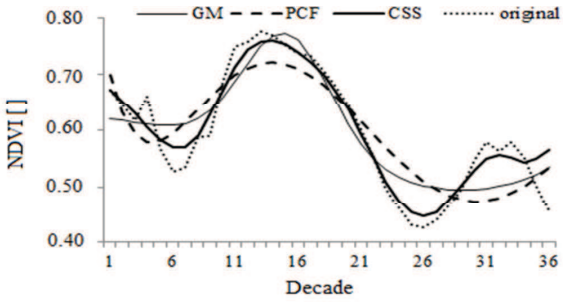


Fig. 3. CDF original NDVI data and PhenoSat fitting results obtained using GM, PCF and CSS algorithms.

Table II. PhenoSat estimated phenological dates (10-days) obtained for Closed Deciduous Forest (CDF).

NDVI metrics	CDF					
	DL	SG	CSS	PCF	GM	FS
Gup	9	7	7	3	9	7
LIP	10	10	10	8	12	10
Mat	11	11	11	13	14	12
Max	14	13	14	14	15	13
Se	18	17	19	14	16	14
RIP	21	22	21	22	19	22
Dm	24	24	23.5	26	21	25
R-square	0.98	0.93	0.93	0.71	0.80	0.91

The results for the VIN (Table III) showed PCF and FS with earlier green-up occurrences, comparing with the other algorithms. For FS and GM the senescence and maximum stages occurred on the same day of the year. PCF had the lower r-square (0.23) value, followed by GM (0.53) and FS (0.57). The algorithms with higher r-square were the DL, SG and CSS, with 0.67, 0.66 and 0.66, respectively.

3.2. Comparison with ground-based measures

Vineyard green-up and maximum NDVI metrics obtained by PhenoSat were compared with the budbreak and veraison stages observed in the field for the period 2000-2005.

The CSS and FS algorithms were unable to detect with precision the maximum stage. However, for the green-up they presented correlations of 0.59 and 0.60. The opposite was verified with PCF, which obtained correlations of 0.62 and 0.15 for the maximum and green-up phenological stages, respectively. DL presented a more coherent behavior between the two phenological stages, being the best algorithm with both correlations above 0.70.

The PCF, GM, FS and CSS results were obtained using smoothing parameters of 4, 4, 4 and 0.4, respectively. The sensitivity of each algorithm to the smoothing parameter was tested running PhenoSat 8 times for each algorithm. For PCF, GM and FS the parameter was varied between 1 and 8, and for CSS was varied from 0.1 to 0.8. For the green-up

Table III. PhenoSat estimated phenological dates (10-days) obtained for Deciduous Shrubland with Sparse Trees (DSST) and Vineyard (VIN), for the year 2000.

NDVI Metrics	DSST						VIN					
	DL	SG	CSS	PCF	GM	FS	DL	SG	CSS	PCF	GM	FS
Gup	13	13	13	9	14	10	12	12	12	9	12	11
LIP	17	16	16	16	19	17	14	14	14	14	15	15
Mat	20	17	17	24	25	22	16	16	15	20	18	18
Max	26	26	26	25	26	26	18	20	19	21	19	19
Se	29	34	34	25	26	30	20	24	23	22	19	19
RIP	31	35	35	34	33	33	22	28	24	28	21	22
Dm	33	36	36	36	36	36	26	34	31	33	23	33
R-square	0.99	0.99	0.99	0.98	0.96	0.99	0.67	0.66	0.66	0.23	0.53	0.57

stage, the results showed that higher parameters obtained higher correlations between estimated and ground-based observations (Table IV). For the maximum stage, intermediate values (4, 5 and 6) obtained the best results except for CSS, where 0.1 is the smoothing parameter with the highest correlation (0.40).

Table IV. Spearman correlation rank (rho) between PhenoSat estimated phenological parameters and ground-based measures, for a Vineyard region from 2000 to 2005.

Algorithm	*Gup*			*Max*		
	Par	rho	p	Par	rho	p
CSS	0.8	-0.71	0.06	0.1	-0.40	0.22
DL	—	-0.77	0.04	—	0.71	0.06
FS	8	-0.83	0.02	6	-0.43	0.20
GM	8	-0.89	0.01	5	-0.60	0.10
PCF	8	-0.64	0.09	4	0.62	0.10
SG	—	-0.15	0.39	—	0.15	0.39

Par is the best smoothing parameter and p represents the p-value.

4. CONCLUSIONS

The experiment carried out showed that PhenoSat is capable to extract phenological information from satellite vegetation data. All six fitting algorithms tested proved to be capable to detect the peaks and transitions of the main growing season with high precision. However, the PCF and GM algorithms presented the most distinct results being unable to accurately capture the double growing season, over smoothing this period.

Comparing ground-based measures and PhenoSat derived phenology, the DL presented correlations above 0.70 for both phenological stages evaluated. For CSS, FS, GM and PCF algorithms, high smoothing parameters allowed a more accurate detection of the green-up occurrence, presenting correlations of 0.71, 0.83, 0.89 and 0.64, respectively. However, for the maximum stage, intermediate values obtained the best correlations except for CSS, where the lowest parameter (0.1) produced the highest correlation (0.40).

A preliminary version of PhenoSat is freely available at http://www.fc.up.pt/LamSat_XXI.

5. REFERENCES

[1] A. Rodrigues, A. R. S. Marçal and M. Cunha, "PhenoSat- A tool for vegetation temporal analysis from satellite image data", *6th international Workshop on the Analysis of Multi-Temporal Remote Sensing Images, Multi-Temp 2011*, art. no. 6005044, IEEE Xplore digital library 978-1-4577-1203-6/11:45-48, 2011.

[2] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Savitzky-Golay Smoothing Filters", *Numerical Recipes: The Art of Scientific Computing (3rd ed.),* New York: Cambridge University Press, ISBN 978-0-521-88068-8, 2007.

[3] X. Zhang, M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. G. F. Hodges, F. Gao, B. C. Reed and A. R. Huete, "Monitoring vegetation phenology using MODIS", *Remote Sensing of Environment,* vol. 84, no. 3, pp. 471-475, 2003.

[4] M. Cunha, A. R. S. Marçal and A. Rodrigues, "A Comparative Study of Satellite and Ground-Based Vineyard Phenology", *Proc. 29th EARSeL Symp., Imagin[e,g] Europe*, pp. 68-77, 2010.

[5] A. Fischer, "A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters", *Remote Sensing of Environment,* vol. 48, no. 2, pp. 220-230, 1994.

[6] C. H. Reinsch, "Smoothing by Spline Functions". *Numerische Mathematik,* vol. 10, pp. 177-183, 1967.

[7] J. Verschelde, "MCS320: Introduction to Symbolic Computation: Matlab Lecture 3: Polynomials and Curve Fitting", pp. 1-3, 2007.

[8] MathWorks, "Curve Fitting Toolbox: Fit curves and surfaces to data using regression, interpolation, and smoothing", 2011.

[9] A. Goshtasby and W. D. O´Neill, "Curve fitting by a Sum of Gaussians", *Graphical Models and Image Processing,* vol. 56, no. 4, pp. 281-288, 1994.

APPENDIX 3

PhenoSat – A tool for remote sensing based analysis of vegetation dynamics

PhenoSat – A tool for remote sensing based analysis of vegetation dynamics

Arlete Rodrigues[*1], André R.S. Marcal[2] and Mário Cunha[1]

[1] Faculdade de Ciências da Universidade do Porto, Dept. GAOT & Centro de Investigação em Ciências Geo-Espaciais, Porto, Portugal (dr.arlete@gmail.com; mcunha@mail.icav.iup.pt)

[2] Faculdade de Ciências da Universidade do Porto, Dept. Mat. & INESC TEC, Porto, Portugal (andre.marcal@fc.up.pt)

Abstract PhenoSat is a software tool that extracts phenological information from satellite based vegetation index time-series. This paper presents PhenoSat and tests its main characteristics and functionalities using a multi-year experiment and different vegetation types – vineyard and semi-natural meadows. Three important features were analyzed: (1) the extraction of phenological information for the main growing season, (2) detection and estimation of double growth season parameters, and (3) the advantages of selecting a sub-temporal region of interest. Temporal NDVI satellite data from SPOT VEGETATION and NOAA AVHRR were used. Six fitting methods were applied to filter the satellite noise data: cubic splines, piecewise-logistic, Gaussian models, Fourier series, polynomial curve-fitting and Savitzky-Golay. PhenoSat showed to be capable to extract phenological information consistent with field measurements, presenting in some cases correlations above 70% (n=10; $p \leq 0.012$). The start of in-season regrowth in semi-natural meadows was detected with a precision lower than 10-days. The selection of a temporal region of interest, improve the fitting process (R-square increased from 0.596 to 0.997). This improvement detected more accurately the maximum vegetation development and provided more reliable results. PhenoSat showed to be capable to adapt to different vegetation types, and different satellite data sources, proving to be an useful tool to extract metrics related with vegetation dynamics.

1 Introduction

Temporal vegetation profiles based on remotely sensed data provide valuable information for understanding land cover dynamics, generally interpreted by vegetation phenological events. Sensors such as Advanced Very High Resolution Radiometer (AVHRR), SPOT VEGETATION (Satellite Pour l' Observation de la Terre - Vegetation), MODIS (Moderate-Resolution Imaging Spectroradiometer), MERIS (Medium Resolution Imaging Spectrometer) and PROBA-V (Project for On-Board Autonomy - Vegetation) are capable to acquire appropriate temporal vegetation index (VI) that could be related with vegetation phenology. The analysis of the VI time-series through time allows the extraction of appropriate metrics to describe vegetation dynamics (Bradley and Mustard 2008). Although the access to satellite vegetation data is currently widespread, with low or no costs, there is still a gap between the data and meaningful information. The large amounts of data and the presence of noise can make the analysis and extraction of relevant vegetation information a difficult and time consuming process.

The VI time-series obtained by Earth Observation Satellite (EOS) images generally include various noise components such as atmospheric disturbances, viewing and solar illumination variability, cloud cover and others. Maximum Value Composites (MVC) are used to minimize the noise influence by analyzing the VI values on a pixel-by-pixel basis, in a predefined time-period, retaining the highest value for each pixel location (Holben 2007). A MVC image is obtained when all pixels have been evaluated. The MVC imagery is highly related to the green vegetation dynamics, and common problems encountered in single-date remote sensing studies, as cloud contamination, atmospheric attenuation and observation geometry, are minimized using MVC (Tucker et al. 1985). However, generally the MVC process is not sufficient to eliminate all unrealistic variability from VI time-series (Jonsson and Eklundh 2004; Rodrigues et al. 2013). Furthermore, additional noise may be also introduced by the process of overlaying several images (for example due to image misregistration). Thus, it is necessary to fit a model to the observed data before the extraction of vegetation dynamics information. An appropriate model should be capable of smoothing the data without

[*] Corresponding author: Email: dr.arlete@gmail.com, Tel. (+351)220402160

introducing artifacts or suppressing natural variations of the vegetation (Fontana et al. 2008). Over recent years, a large number of fitting algorithms have been proposed for this purpose, including the ARMD3-ARMAS, Fourier, asymmetric Gaussian, Savitzky-Golay, double logistic, and mean-value iteration, to name a few (Fillipova-Racheva and Hall-Beyer 2000; Roerink at al. 2000; Jonsson and Eklundh 2002; Chen et al. 2004; Beck et al. 2006; Ma and Veroustraete 2006).

A number of software tools able to extract phenology from satellite VI time-series have been developed, such as TIMESAT (Jonsson and Eklundh 2004), TimeStats (Udelhoven 2011), Enhanced TIMESAT (Tan et al. 2011), PPET (McKellip et al. 2010), and the software developed by USGS Earth Resources Observation and Science Center (Ross et al. 2009). Although these software products have important functionalities for the extraction of phenological information, they present some limitations. They include: i) none of them allows the selection of an in-season window of interest, which is fundamental for analyzing vegetation types and crop systems with more than one growth cycle through the year; and ii) except for TIMESAT, none of them has a specific option to characterize a double growth season phenology. Moreover, the software packages were tested only in short number of vegetation types and detailed comparisons with field data are still scarce. To address these limitations, PhenoSat was developed to detect the number of growth seasons in each year and has the option to define an in-season window of interest. The main characteristics and functionalities of PhenoSat were tested using a multi-year experiment and different vegetation types, as well as data from two different sensors.

2 PhenoSat Description

PhenoSat is a software tool developed to extract phenological information from satellite VI time-series. This tool was implemented in Matlab (Higham and Higham 2000) using a simple interface to provide an easy-to-use software. PhenoSat can receive two standard input text files: containing the original satellite VI images, or containing a temporal VI dataset. For the VI images, a pixel-by-pixel approach is conducted, and a specific region can be selected instead of using all image size. For a temporal dataset, the numerical values (VI) are already standardized in a text file.

A number of satellite based metrics related with main growing season phenological parameters (e.g. start of season, maximum vegetation development, end of season) can be determined by PhenoSat. Some vegetation types and crops systems present more than one growth cycle through the year manly related with crop rotation or vegetation regrowth. The timing and magnitude of these in-season cycles present high intra-annual variability due to some factors such as climate, animal grazing and human land use decisions. Information about the timing of start and maximum of these seasonal cycles can be obtained using PhenoSat. It is important to note that some extreme conditions (e.g. fire, unseasonal snow) could result in a false report of a double growing season. For this reason, a new feature was developed in PhenoSat that allows the selection of a sub-temporal region of interest, based on vegetation dynamics knowledge. The annual VI time-series subinterval, defined manually or automatically, improves the fitting process, providing more realistic results of the vegetation dynamics.

PhenoSat outputs two files containing the phenological information (estimated date and respective VI value) and the processed data at each of the fitting steps. When the VI images are used as input, three additional output images will be created. These images present, for each pixel analyzed, the phenological estimated dates for three main stages: start of season, maximum vegetation development and end of season.

2.1 PhenoSat Fitting Methods

Some VI datasets available online from different sensors (e.g. AVHRR, SPOT VEGETATION (SPOT_VGT), MODIS) are already preprocessed in order to reduce the noise caused by a variety of bio-physical factors (Carreiras et al. 2003; Gutman 1991; Li and Strahler 1992). Although this preprocessing is generally effective, the VI datasets still retain some problems (punctual outliers or abrupt changes) that require additional processing. Noise reduction filters can be applied to remove the undesirable artifacts, improving the subsequent analysis, and leading to more reliable vegetation dynamics information.

PhenoSat considers as outliers the VI values that present a VI difference above 0.2 from the median (Mw), and from its left and right spatial neighbors. The values of these outliers are replaced by the Mw value. To enhance the spring and summer periods, an upper envelope (Beck et al. 2006) can be applied, allowing a better discrimination of the maximum vegetation development. Although these actions can remove the VI time-series outliers, some noise might still remain. For this reason, PhenoSat provides six methods that can be used to obtain improvements in the noise reduction process. The methods are: cubic smoothing splines (CSS), piecewise-logistic (PL), Gaussian models (GM), Fourier series (FS), polynomial curve-fitting (PCF) and Savitzky-Golay (SG). The CSS algorithm (Reinsch 1967) fits a cubic smoothing spline to the VI time-series data. The adherence of the smoothing spline method to the given data depends on the algorithm parameter selected.

Beck et al. (2006) and Fontana et al. (2008) proved that vegetation dynamics tends to follow a well-defined growth temporal pattern and the vegetation cycle can be represented by a double-logistic function. PhenoSat uses a particular case of a double-logistic function (PL) defined by Eq. 16.1, where t represents the time, VI_t the VI value at time t, c and d are the slopes at the 'left' and the 'right', and p and e are the inflection points dates. VI_w and VI_{w1} are the VI values before the bud break and after the leaf fall, respectively. The k parameter is related with the asymptotical value and assures the continuity between vegetation growth and senescence parts, even when they differ in shape (Cunha et al. 2010). The PL seven parameters are estimated using the Levenberg-Marquardt algorithm (Montgomery et al. 2006), which requires reasonable initial values.

$$VI_t = VI_w + \frac{k}{1+\exp[-c(t-p)]} - \frac{k+VI_w-VI_{w1}}{1+\exp[-d(t-e)]} \qquad (16.1)$$

Figure 16.1 presents a representation of the PL parameters, using two consecutive years of NDVI (Normalized Difference Vegetation Index) SPOT_VGT data. The continuity between the two years is assured by the VI_w and VI_{w1}, being the VI_w for the second year (beginning of the time-series) the same as the VI_{w1} for the first year (final of the time-series).

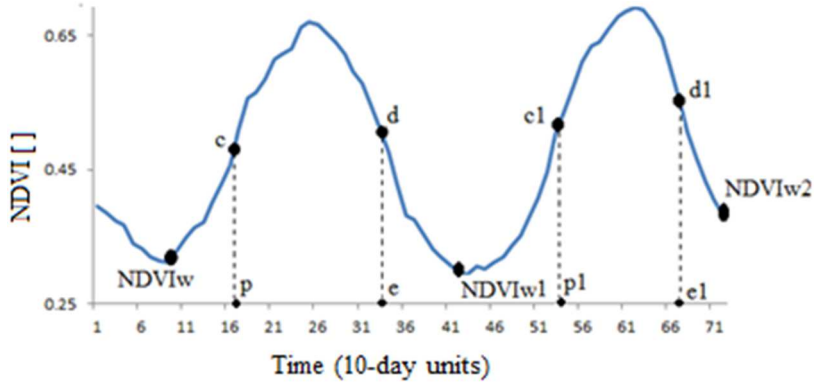


Fig. 16.1 Representation of the piecewise-logistic function parameters, using NDVI SPOT VEGETATION data from two consecutive years.

The GM adjustment to data is given by the Eq. 16.2 (Goshtasby and Oneill 1994):

$$y = \sum_{i=1}^{n} a_i e^{[-(\frac{x-b_i}{c_i})^2]} \qquad (16.2)$$

where a is the amplitude, b the centroid (location), c is related to the peak width and n is the number of peaks to fit ($1 \leq n \leq 8$).

The FS is a sum of sine and cosine functions of different period that describes a periodic signal (Mitra 2010). In the trigonometric form, it is represented as:

$$y = a_0 + \sum_{i=1}^{n} [a_i \cos(nwx) + b_i sin(nwx)] \qquad (16.3)$$

where a_0 models a constant (intercept) term in the data and is associated with the $i = 0$ cosine term, w is the fundamental frequency of the signal, and n is the number of terms (harmonics) in the series ($1 \leq n \leq 8$).

The PCF (Verschelde 2007) finds the coefficients of a polynomial, of a given degree, that fits the data. The higher the degree, the closer the fitting curve will be to the data, although this should be done only up to a certain (reasonable) degree.

The SG filter (Press et al. 2007) defined by

$$g_i = \sum_{nL}^{nR} c_n f_i + n \qquad (16.4)$$

is a particular type of low-pass filter, that replaces each data value $f_i,\ i = 1, \ldots, N$, by a linear combination g_i of nearby values in a window defined by the number of points 'to the left' (nL) and 'to the right' (nR) of a data point i. In PhenoSat, the SG filter uses $nL = nR$ and is always applied to smooth the original VI data. For the subsequent analysis, it can be used alone or combined with one of the other methods.

Figure 16.2 presents a comparison of the six fitting methods described, using a NDVI SPOT_VGT annual time-series obtained from a Closed Deciduous Forest. The PCF and GM methods present the most distinct fitting results for the main and double growing seasons. The biggest differences are on the double

growth season, where these two methods present low sensitivity to detect the regrowth peak, over smoothing this occurrence.

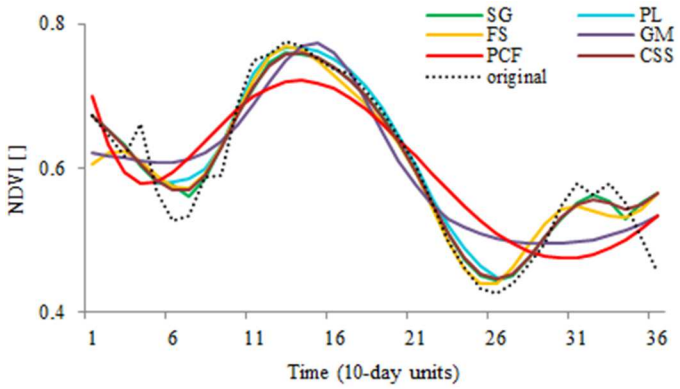


Fig. 16.2 Comparison of PhenoSat fitting methods using a NDVI SPOT VEGETATION annual time-series obtained from a Closed Deciduous Forest.

2.2 Phenological Metrics

PhenoSat is able to determine the VI value and time of occurrence of the following seven phenological events in the main growing season: start of season (SOS), maturity (MAT- beginning of the ripening stage/full canopy), maximum vegetation development (MVD), senescence (SEN), end of season (EOS), and the maximum growth and maximum senescence rates (namely left (LIP) and right (RIP) inflexion points, respectively).

The phenological information is obtained using the derivatives of the fitting VI time-series, as illustrated in figure 16.3. The LIP (or RIP) corresponds to the maximum (or minimum) of the fitted first derivative. The MVD is determined as the maximum VI fitted value. The maxima of the fitted VI time-series second derivative, at the left/right of the MVD, identify the SOS/EOS. Similarly, the MAT/SEN can be found using the minima of the fitted data second derivative, at the left/right of the MVD.

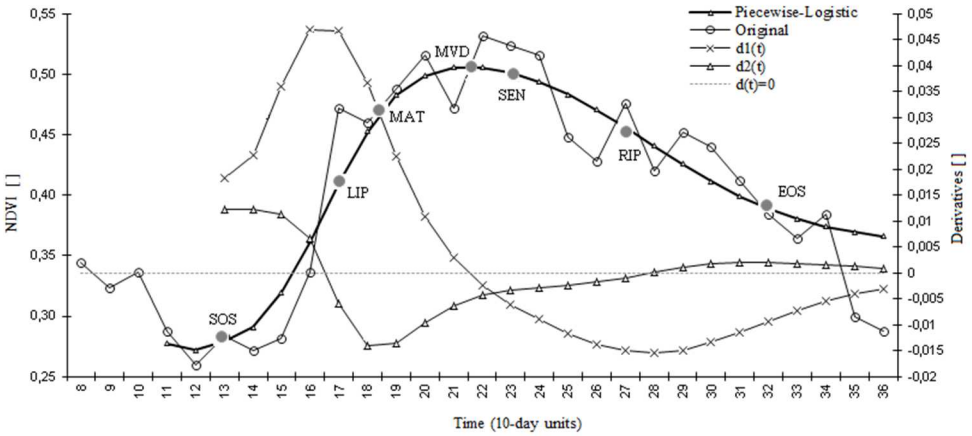


Fig. 16.3 Representation of PhenoSat derived phenological stages using the maxima and minima of the first and second derivatives of the NDVI fitted data.

Some factors, such as adverse weather conditions (snow, ice or extreme heat), water availability, pasture management and/or herbaceous vegetation growth in the winter season, can induce an annual regrowth in some crops (e.g. crop systems with more than one growth cycle a year, shrublands or semi-natural meadows). This phenological information can also be extracted by PhenoSat if required. This option allows to record the VI value and date of occurrence for the start and maximum of this in-season period. PhenoSat calculates the regrowth parameters using the smoothed time-series (before the application of the fitting method) after the EOS time occurrence.

The regrowth start is defined as the point where an increase of three or more points occurs after the EOS stage. After this starting point, a decreasing period of two or more points determines the maximum of the regrowth. In some cases, the regrowth reported can be a "false regrowth". For example in vineyards, as many other discontinuous canopies, during the winter season the inter-row vegetation growth appears as a regrowth in the vineyard annual profile. The unseasonal snow could also result in a false report of a double growth season in many environments. Only with the knowledge/analysis of the ground conditions it is possible to infer about the truth of the regrowth.

The selection of an in-season temporal region of interest, based on known vegetation dynamics, can help dealing with a false regrowth, particularly for natural land cover types. PhenoSat has the possibility to select, automatically or manually, the in-season temporal region of interest. The manual selection can be done by inputting the initial and final time positions, based on known behavior of the vegetation in the field at normal growth conditions. This type of selection presents some limitations when adapting to different dynamics over the years. PhenoSat tries to solve this problem with an optional approach that automatically detects the region of interest. This option, based on the VI time-series profiles, is more flexible and can adapt to the dynamics variations over the years. To determine the annual time-series subinterval, PhenoSat firstly calculates the maximum value of the VI data. Then, searches for the initial position, which corresponds to the point where a significant increase (or abrupt decrease) is verified to the left of the maximum. Afterwards, to determine the final position, the algorithm proceeds in a similar way but evaluating the data to the right of the maximum.

3 PhenoSat Application

3.1 Study Area and Satellite Data

The PhenoSat software was tested for different vegetation types and geographical locations in continental Portugal. The NDVI time-series from AVHRR (10-days composite and 1-km resolution) and SPOT_VGT (10-days composite and 1-km resolution) covering Portugal, were downloaded from The Joint Research Centre Community Image Data portal (JRC-CID 2013).

The performances of PhenoSat were tested in two land use types that present, mainly, a different annual growth pattern (Table 16.1): vineyard (VIN) in Douro wine region (Northeast Portugal); and semi-natural meadows (SNM) in Montalegre (Northeast Portugal). In Douro region the predominant land cover is the vineyard with extensive contiguous areas. The vineyard has a long dormancy period with intense understory vegetation growth and a discontinuous canopy (Cunha et al 2010). The SNM are an essential element of the mountain landscapes in Northern Portugal, and represent the main fodder resource for livestock production. This type of crop is characterized by a regrowth around the month of August, whose intensity and date of occurrence are mainly dependent of climatic conditions (Pocas et al. 2012).

The different vegetation profiles provided by these crops (Fig. 16.4) permit to evaluate the adaptability and performance of PhenoSat to distinct situations. For each crop, a training area was defined carefully to avoid pixel boundaries with other crops, hence the reduced number of pixels (Table 16.1) used in these experiments. Entire training areas were considered as units, instead of using a pixel-by-pixel approach. The

median of the NDVI values of the pixels assigned to each crop were computed. The yearly NDVI time-series were created using the median values obtained for each of the images available in a year.

Table 16.1 Description of training areas and satellite datasets used to test PhenoSat

Land Cover	Acronym	Coordinates (Long/Lat WGS84)	Satellite products	Time-series period	Size (pixels)
Vineyard *	VIN	UL: 7d45'17.7W, 41d09'51.6N BR: 7d43'41.8W, 41d08'48.4N	SPOT_VGT AVHRR	2001-2010	6
Semi-Natural Meadows	SNM	UL: 7d57'36.9W, 41d38'15.2N BR: 7d56'33.4W, 41d37'11.6N	SPOT_VGT AVHRR	2001-2010	4

* phenological field measures available

UL: upper left corner; BR: bottom right corner

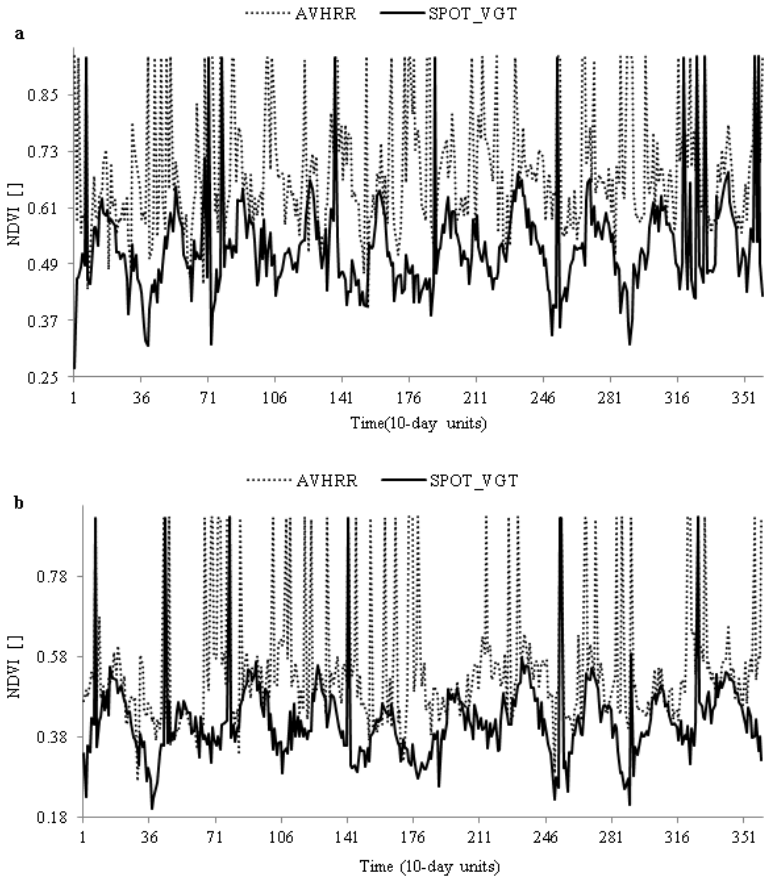


Fig. 16.4 Original NDVI temporal profiles obtained from SPOT VEGETATION and AVHRR data, for the semi-natural meadows (**a**) and vineyard (**b**) land use types, for the period 2001-2010.

3.2 Extraction of Phenological Information

The ability of PhenoSat to estimate phenological metrics from satellite VI data was evaluated by a comparison between PhenoSat derived phenology and field measures. Table 16.2 presents the statistics of field phenological measures obtained for each study area. For the VIN test site, phenological measures collected in the field according to the Baggiolini scale (Baggiolini 1952) were available. The bud break (BUB), flowering (FLO) and veraison (VER, define as the 'change of color grapes' stage) field observations were compared with the SOS, MAT and SEN derived by PhenoSat. As no ground measures of phenology were available for SNM, the PhenoSat results for SNM were compared with the reference measures (named field measures from this point) derived by visual inspection of the original VI time-series, taking into account the knowledge of the vegetation behavior in the field at normal conditions. As an example, figure 16.5 presents the field measures determined from the SNM for one year. The SOS was determined as the first point where a significant (four or more points) NDVI growth was occurred (March/April). The MVD was identified as the maximum NDVI value in the annual time-series, which generally occurs in June or early July. The abrupt decrease verified after this point is due to the grass cutting process. The remaining ground vegetation (about 5cm height) begins a senescence period until the maximum senescence (EOS), occurring mostly around August. In general the SNM EOS stage is followed by a regrowth (RG), representing the first significant (three or more points) vegetation growth after the EOS occurrence.

Table 16.2 Statistics of field phenological measures obtained for vineyard (VIN) and semi-natural meadows (SNM) vegetation types

Phenological Stage	Statistics	VIN	SNM
Start of Season	Mean (DOY)	82.16	98.00
	Maximum (DOY)	92.00	150.00
	Minimum (DOY)	78.00	70.00
	Standard deviation (days)	4.73	23.15
Flowering	Mean (DOY)	145.76	----
	Maximum (DOY)	153.00	----
	Minimum (DOY)	125.30	----
	Standard deviation (days)	7.18	----
Veraison / Maximum Vegetation Development	Mean (DOY)	204.20	169.00
	Maximum (DOY)	213.50	180.00
	Minimum (DOY)	199.70	150.00
	Standard deviation (days)	4.07	10.44
End of season	Mean (DOY)	----	237.00
	Maximum (DOY)	----	250.00
	Minimum (DOY)	----	210.00
	Standard deviation (days)	----	12.69

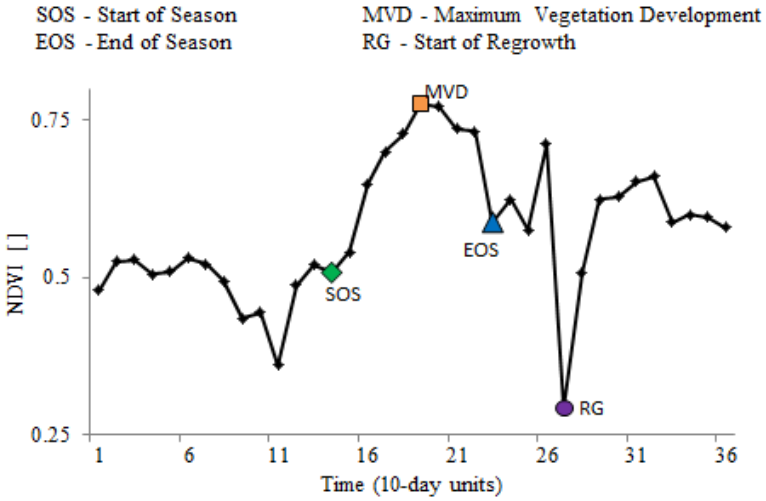


Fig. 16.5 Example of phenological "reference measures" derived from the analysis of a 1-year series (from NDVI SPOT VEGETATION for the semi-natural meadows).

4 Results and Discussion

4.1 PhenoSat-derived phenology

The VIN phenological parameters estimated by PhenoSat were compared with those obtained from field measurements. The results are presented in Table 16.3. Using the SPOT_VGT data, the correlations obtained for the start of season were no higher than 33%, except for GM and PL which obtained, respectively, values of 0.77 (n=10; p=0.004) and 0.63 (n=10; p=0.025). The inter-row vegetation growth during the winter, and the difficulty in discriminating the first grapevine leaves from the satellite data, makes it difficult to estimate with high precision the SOS VIN stage. For maturity and mid-season, the correlations were, in general, higher than for the SOS. The maturity phenological stage represents the period of VIN full canopy which can be better identified and determined using NDVI data, as it is showed by the correlation values.

The NDVI values for AVHRR were always greater than SPOT_VGT values (Fig. 16.4**b**), presenting the higher differences at the end of the years. Comparing with field measurements (Table 16.3), the AVHRR data achieved better correlations for the start of season, for most of the fitting methods. However, for the remaining parameters, the SPOT_VGT data were better, providing correlations above 70% (n=10; $p \leq 0.012$), in some cases.

The flexibility of PhenoSat to extract phenology data from different land use types was tested using the SNM. A comparison between the estimations and field measures is presented in Table 16.4. The phenological dates for SOS, MVD and EOS stages were extracted with a reasonable precision with correlations higher than 0.50 in most cases. All the fitting methods produced similar results, being PCF the method with best performance for the SOS stage (n=10; r=0.86; p<0.001).

The capability of PhenoSat in determining multiple growths in a same year was also tested using the SNM data. This crop is characterized by an annual regrowth around the month of August. However, the start of the regrowth can suffer changes due to some factors such as the climatic conditions and human intervention.

Table 16.3 Correlations between field and PhenoSat estimated vineyard phenology, using NDVI data from SPOT_VGT and AVHRR sensors

Fitting Methods	SPOT_VGT (n=10)			AVHRR (n=10)		
	Start	Maturity	Mid-season	Start	Maturity	Mid-season
CSS	-0.27	-0.78	-0.40	0.49	-0.28	0.07
FS	-0.17	-0.76	-0.58	0.56	-0.32	-0.27
GM	0.77	-0.69	0.67	0.45	0.47	0.36
PCF	-0.33	-0.71	-0.38	-0.30	0.41	0.21
PL	0.63	-0.66	-0.55	-0.06	0.30	0.18
SG	0.30	-0.77	-0.25	0.30	-0.21	-0.08
Mean (absolut)	0.37	0.73	0.47	0.36	0.30	0.20

The start, maturity and mid-season represent the comparison between SOSvsBUB, MATvsFLO and SENvsVER, respectively. The SOS, MAT and SEN are the derived PhenoSat phenology, and BUB, FLO and VER are the phenological measures obtained in the vineyard
Fitting methods: CSS – Cubic Smoothing Splines; FS – Fourier Series; GM – Gaussian Models; PCF – Polynomial Curve Fitting; PL – Piecewise-Logistic; SG – Savitzky-Golay

Table 16.4 Comparison between field and PhenoSat estimated phenology for the semi-natural meadows crop, using SPOT_VGT data

	Correlation Field vs PhenoSat		
Method	Start of Season	Maximum Vegetation Development	End of Season
CSS	0.58	0.54	0.51
FS	0.50	0.45	0.66
GM	0.43	0.56	0.63
PCF	0.86	0.44	0.63
PL	0.38	0.51	0.65
SG	0.53	0.53	0.54

Figure 16.6 presents the original and smoothed NDVI SPOT_VGT profiles for the SNM land use type, for the three final years (2008, 2009 and 2010). The smoothed data were obtained using a SG filter with a first degree polynomial and frame size 5. These parameters removed the outliers/spikes without suppressing the natural variations of the SNM VI original data. From the analysis of the smoothed profiles it is possible to see that 2010 is the only year that presents a double growth season, with start (3 or more consecutive points increasing) around the DOY 270. Table 16.5 shows the timing of regrowth derived from the original data and determined using PhenoSat. All six fitting methods were capable to detect the start of the regrowth, obtaining similar results. In the years 2001, 2003, 2007, 2008 and 2009 there was no regrowth, which was correctly verified by PhenoSat. For the remaining years, PhenoSat accurately detected the beginning of the double growth season, being the differences between original and estimated parameters of 10-days (except for PCF in the year 2004). Similar conclusions were observed for the maximum of the regrowth.

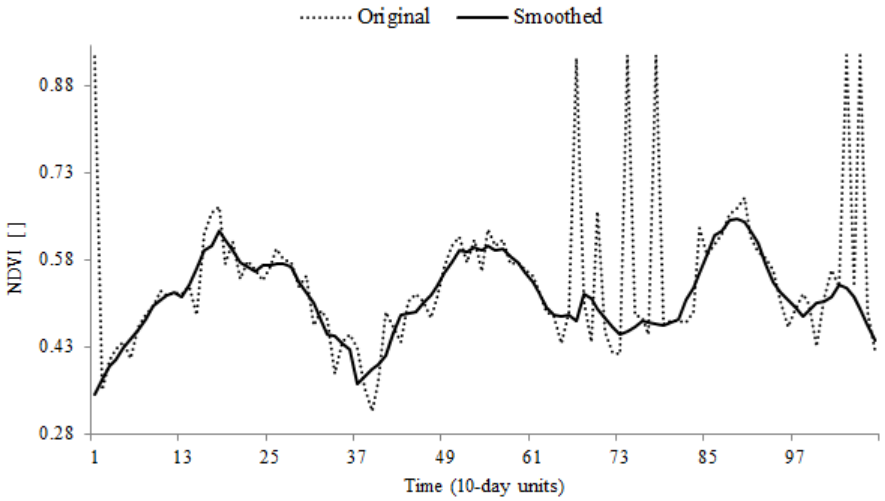


Fig. 16.6 Original and smoothed NDVI SPOT VEGETATION time-series for the semi-natural meadows crop, for years 2008, 2009 and 2010.

Table 16.5 Start of double growth season estimations using original and fitted data, for semi-natural meadows

	Start of double growth (day of the year)									
Method	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
CSS	n.a.	260	n.a.	230	260	250	n.a.	n.a	n.a.	260
FS	n.a.	270	n.a.	230	260	260	n.a.	n.a.	n.a.	260
GM	n.a.	260	n.a.	230	260	260	n.a.	n.a.	n.a.	270
PCF	n.a.	260	n.a.	250	270	260	n.a.	n.a.	n.a.	270
PL	n.a.	260	n.a.	230	260	250	n.a.	n.a.	n.a.	260
SG	n.a.	260	n.a.	230	260	250	n.a.	n.a.	n.a.	280
Original	n.a.	260	n.a.	230	260	250	n.a.	n.a.	n.a.	270

n.a. signifies that no regrowth is verified on this year

4.2 Advantages of selecting an in-season region of interest

PhenoSat has the option to select a region of interest, instead of using all range of observations in a year. The reduction of the VI time-series improves the fitting process, capturing more efficiently the maximum vegetation development, thus producing more realistic results. To evaluate the utility of this feature on phenological studies, PhenoSat was tested using the VIN. The interest region must be selected according to the behavior of the studied vegetation in the field, under normal conditions. The grape-growth cycle in Douro (Portugal) starts with the bud break stage, which occurs around March. The harvest period typically occurs between August and September/October, however it is deeply dependent on the winemakers according to the style and quality of the wine they wish to produce. Considering these facts, the main phenological cycle of the studied VIN crop is assumed to be ranged from March (DOY 70) to September (DOY 270).

Figure 16.7 presents the VIN NDVI SPOT_VGT data for the year 2009, and the PL fitting results using all range of observations (dash line) and the in-season region of interest. Using all range of observations, the maximum peak of the VIN (around the DOY 180/190) cannot be detected due to the initial peak around the DOY 30 that could be related with winter vegetation growth in the vineyard inter-row. The inclusion of

this early pick of NDVI profile led to an over smoothed of the main growing cycle. On the other hand, the use of the region of interest allowed a more accurate adaptation of the fitting method to the variations of the original data during the main growth cycle. The full canopy and senescence stages were captured with high precision and more realistic results were produced. The PL fitted data, produced using region of interest (from 7 to 27 ten days NDVI; 21 observations) instead the all 36 observations, improves the R-square from 0.596 to 0.997.

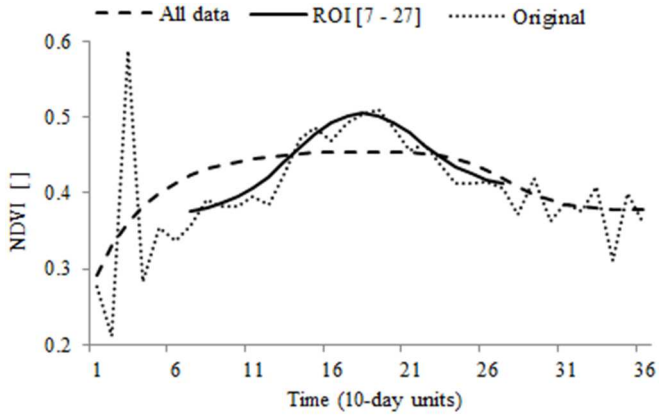


Fig. 16.7 NDVI SPOT VEGETATION original time-series (dot line) and the piecewise-logistic fitted results using all range of observations (dash line) and the in-season region of interest (solid line), for the vineyard in 2009

Another example of the importance of the use of the interest zone in vineyards is presented in figure 16.8. The PL fitted results, using all the 36 observations, captured the initial peak (DOY 70) as the maximum development of the VIN crop. This erroneous information led to non-accurate phenological estimations. Using the region of interest, the fitted results captured more precisely the VIN growing season, over smoothing the period related with the soil vegetation growth.

The selection of a region of interest proved to be useful not only in reducing the processing time, but also in obtaining better fitted results, and consequently more reliable phenological information.

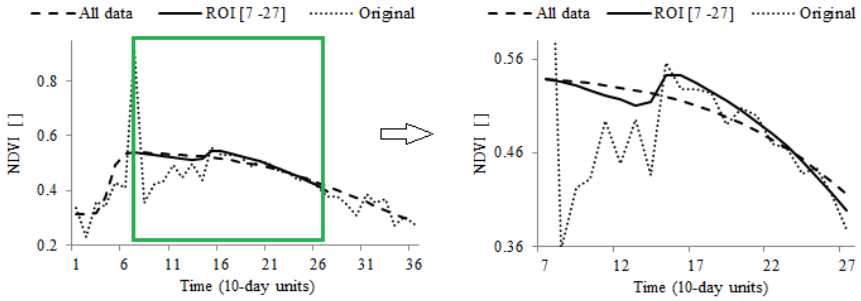


Fig. 16.8 NDVI SPOT VEGETATION original time-series (dot line) and the piecewise-logistic fitted results using all range of observations (dash line) and using an in-season region of interest (solid line), for the vineyard in 2001

5 Conclusions

PhenoSat is an easy-to-use software tool which enables phenological information to be extracted from satellite VI data. The experiments carried out indicate that the phenological estimations provided by PhenoSat are accurate and consistent with field measurements.

PhenoSat permits the detection of an annual regrowth and the possibility to define an in-season region of interest, which are limitations of other software packages used to extract phenology.

The option to select an in-season region of interest results on an improvement of the fitting process, leading to more reliable results. This PhenoSat feature proved to be a valuable tool for vineyard monitoring and can extend the PhenoSat application to crops with discontinuous canopy, like forestry and deciduous fruit trees. PhenoSat proved to be capable to detect efficiently the regrowth occurrence. The independency of the fitted results leads to a more realistic time-series profile over the year and, thus, more accurate regrowth-derived results.

Comparing PhenoSat with other tools available for phenological studies (e.g. TIMESAT, TimeStats, Enhanced TIMESAT, PPET), PhenoSat appears as an intuitive, easy-to-use software with two new important features: the possibility to select an in-season region of interest, and the capability of identifying multiple regrowth within a single year. Moreover, the extraction of phenological parameters using an algorithm based on changes of growth rates allows PhenoSat to avoid thresholds or empirical constants, providing a flexible tool that can be applied to different crops and VI data provided from different data sources.

PhenoSat is freely available at http://www.fc.up.pt/PhenoSat website.

Acknowledgments

The authors would like to thank JOINT RESEARCH CENTRE (Community Image Data portal) for providing access to the SPOT_VGT and AVHRR images.

Arlete Rodrigues would like to thank to Fundação para a Ciência e a Tecnologia (FCT) for the Doctoral Grant (SFRH/BD/62189/2009).

References

Baggiolini M (1952) Les stades repères dans le développement annuel de la vigne et leur utilization pratique. Revroman de Agric Vitic Arboric 8:4-6

Beck PSA, Atzberger C, Hogda KA, Johansen B, Skidmore AK (2006) Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. Remote Sens Environ 100(3):321-334

Bradley BA, Mustard JF (2008) Comparison of phenology trends by land cover class: a case study in the Great Basin, USA. Global Change Biol 14:334-346

Carreiras JMB, Pereira JMC, Shimabukuro YE, Stroppiana D (2003) Evaluation of compositing algorithms over the Brazilian Amazon using SPOT-4 Vegetation data. Int J Remote Sens 24(17):3427-3440

Chen J, Jonsson P, Tamura M, Gu Z, Matsushita B, Eklundh L (2004) A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sens Environ 91:332-334

Cunha M, Marcal ARS, Rodrigues A (2010) A comparative study of satellite and ground-based vineyard phenology. In: Proc 29th Symp EARSeL, Chania, Greece: 68-77

Fillipova-Racheva D, Hall_Beyer M (2000) Smoothing of NDVI time series curves for monitoring of vegetation changes in time. In: Ecological Monitoring and Assessment Network National Science Meeting 2000, January 17-22, Toronto, Ontario, Canada.

Fontana F, Rixen C, Jonas T, Aberegg G, Wunderle S (2008) Alpine grassland phenology as seen in AVHRR, VEGETATION and MODIS NDVI time series – A comparison with in situ measurements. Sensors 8(4):2833-2853

Goshtasby A, Oneill WD (1994) Curve-fitting by a sum of Gaussians. CVGIP – Graph. Models Image Process 56(4):281-288

Gutman GG (1991) Vegetation indexes from AVHRR – An update and future-prospects. Remote Sens Environ 35(2/3):121-136

Higham, DJ, Higham NJ (2000) Matlab Guide. SIAM, Philadelphia.

Holben BN (2007) Characteristics of maximum-value composite images from temporal AVHRR data. Int J Remote Sens 7(11):1417-1434

Jonsson P, Eklundh L (2002) Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans Geosci Remote Sens 40(8):1824-1832

Jonsson P, Eklundh L (2004) TIMESAT - a program for analyzing time-series of satellite sensor data. Comput Geosci 30(8):833-845

JRC-CID: Joint Research Centre Community Image Data portal (2013). Available at http://cidportal.jrc.ec.europa.eu/home/ [Accessed 13 July 2013]

Li XW, Strahler AH (1992) Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy-Effect of crown shape and mutual shadowing. IEEE Trans Geosci Remote Sens 30(2):276-292

Ma M, Veroustraete F (2006) Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. Adv Space Res 37:835-840

McKellip RD, Ross KW, Spruce JP, Smoot JC, Ryan RE, Gasser GE, Prados DL, Vaughan RD (2010) Phenological parameters estimation tool. In: NASA Tech. Briefs, September 30, New York.

Mitra SK (2010) Digital Signal Processing: A Computer-Based Approach, 4th ed. McGraw-Hill Science/Engineering/Math (ISBN-13:978-0077366766)

Montgomery D, Peck E, Vining G (2006) Introduction to Linear Regression Analysis, 4th ed. Hoboken, NJ: Wiley

Pocas I, Cunha M, Pereira LS (2012) Dynamics of mountain semi-natural grassland meadows inferred from SPOT-VEGETATION and field spectroradiometer data. Int J RemoteSens 33 (14):4334-4355

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical Recipes: The Art of Scientific Computing, 3rd ed. Cambridge University Press (ISBN-10:0521880688): 766-768

Reinsch CH (1967) Smoothing by spline functions. Numer Math 10:177-183

Rodrigues A, Marcal ARS, Cunha M (2013) Monitoring Vegetation Dynamics Inferred by Satellite Data Using the PhenoSat Tool. IEEE Trans Geosci Remote Sens 51(4):2096-2104

Roerink GJ, Menenti M, Verhoef W (2000) Reconstructing cloud free NDVI composites using Fourier analysis of time series. Int J Remote Sens 21(9):1911-1917

Ross KW, Spiering BA, Kalcic MT (2009) Monitoring phenology as indicator for timing of nutrients inputs in northern gulf watersheds. In: Oceans'09 MTS/IEEE Conference, October 26-29, United States.

Tan B, Morisette JT, Wolfe RE, Gao F, Ederer GA, Nightingale J, Pedelty JA (2011) An Enhanced TIMESAT algorithm for estimation vegetation phenology metrics from MODIS data. IEEE J Sel Topics Appl Earth Observ Remote Sens 4(2):361-371

Tucker CJ, Hielkema JU, Roffey J (1985) The potential of satellite remote sensing of ecological conditions for survey and forecasting desert-locust activity. Int J Remote Sens 6(1):127-138

Udelhoven T (2011) TimeStats: A software tool for the retrieval of temporal patterns from global satellite archives. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 4(2):310-317

Verschelde J (2007) Introduction to Symbolic Computation: MCS320. UIC, Dept of Math, Stat & CS, Springer.

APPENDIX 4

Land cover classification in Rondonia (Amazonia) using NDVI time series data from SPOT-VEGETATION

LAND COVER CLASSIFICATION IN RONDONIA (AMAZONIA) USING NDVI TIME SERIES DATA FROM SPOT-VEGETATION

Deise Furlan[1,2], Arlete Rodrigues[2], Mário Cunha[2*], André R.S. Marcal[2], Maria Victoria Ballester[1]

1. Universidade de São Paulo – USP, Centro de Energia Nuclear na Agricultura – CENA, Piracicaba, Brasil; deisefurlan@cena.usp.br, vicky@cena.usp.br

2. Universidade do Porto, Faculdade de Ciências, Centro de Investigação em Ciências Geo-Espaciais, Porto, Portugal; dr.arlete@gmail.com; andre.marcal@fc.up.pt; mcunha@mail.icav.up.pt

ABSTRACT

Land cover change maps are recognized as an important environmental asset, but are difficult to produce on a routine basis. Earth Observation Satellite (EOS) images have a great potential for land cover mapping, which is mostly based on high resolution image data. However, in tropical areas the use of these images is seriously limited due to presence of clouds. This paper evaluates the ability of temporal based image classification methods to produce land cover maps in tropical regions. A practical test was carried out using VEGETATION satellite data from 1999, 2000 and 2010 for Rondonia (Amazonia), Brazil. A total of 13 land cover classes were identified and grouped hierarchically. Temporal profiles were created for each land cover class, and used as the basis for the classification process, with Jaccard and Dynamic Time Warping used as classifiers. When the training and control samples of the same year were used the classification accuracy is greater than 80% for all tested years. To achieve the accuracy level of 80% when using interannual data for training and control samples, the initial number of classes needs to be aggregated up to 7 classes (1999) or 11 classes (2010).

Keywords: Remote Sensing, SPOT-VEGETATION, GLC2000, Amazonia, time-series

1. INTRODUCTION

The global environmental change research community requires improved and up-to date land cover maps at regional to global scales to support a variety of science and policy applications. This is especially true in Brazilian Amazon, where land cover change induced either by human and natural causes has been unprecedented in recent decades (e.g. Hughton *et al*., 2000, Eva *et al.,* 2004).

While the need for frequent monitoring land cover changes is clear, there is difficult to produce using ground based information alone (e.g Loveland *et al.,* 2000). Hence, the development of new methods based in automatic classification to produce land cover change maps with the range, quality and detail needed by scientists and resource management, is still a huge challenge.

The automatic classification of Earth Observation Satellite (EOS) images is thus frequently used for land cover mapping, mostly based on high resolution image data (e.g. from Landsat or SPOT satellites). However, in tropical areas the presence of cloud cover seriously limits the use of EOS images (Carreiras *et al.,* 2003). In the case of Amazonia, cloud cover is particularly frequent in the rainy season, from October to April. An alternative is to use instead low spatial resolution EOS images, which have a very high acquisition frequency (daily), such as AVHRR (e.g.

Loveland *et al.*, 2000), SPOT-VEGETATION (Carreiras *et al.* 2003) and MODIS (Warlow and Egbert, 2008). This classification approach is slightly different. Instead of using a single multi-spectral image, the temporal yearly NDVI profile is used as a feature vector.

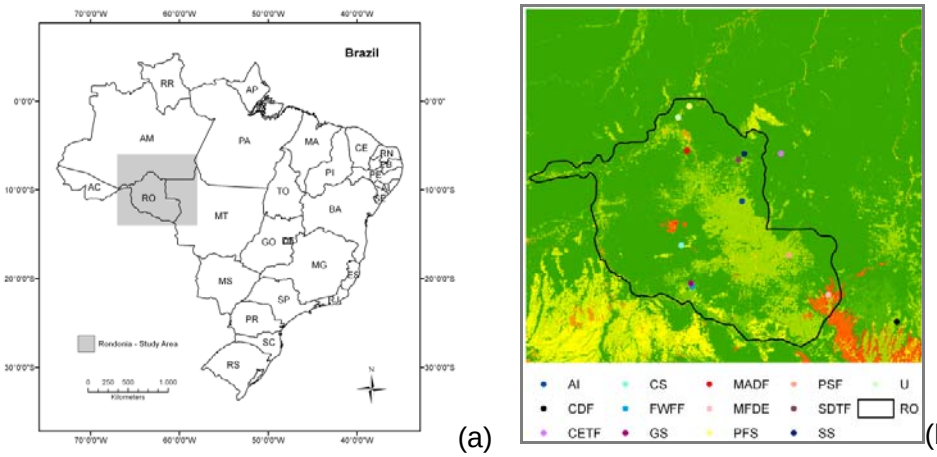
The purpose of this work was to investigate the potential of NDVI temporal series data from SPOT-VEGETATION (VGT) to produce land cover change maps of Rondonia, Amazonia, Brazil.

2. MATERIALS AND METHODS

2.1. Study area

A test area was established in northwest Brazil, corresponding to a rectangular region of 900 by 1000 km, containing the entire state of Rondonia. Figure 1a shows the location of the test area in Brazil.

The Global Land Cover (GLC) 2000 dataset (Eva *et al.,* 2004 was used to establish a number of test sites, characterizing the most relevant land cover classes in this region. A total of 13 land cover classes (LCC) were used: Agriculture Intensive (AI), Closed Deciduous Forest (CDF), Closed Evergreen Tropical Forest (CETF), Closed Shrublands (CS), Fresh Water Flooded Forest (FWFF), Grass Savannah (GS), Mosaic Agriculture / Degraded Forest (MADF), Montana Forest (500-1000m) – Dense Evergreen (MFDE), Periodically Flooded Savannah (PFS), Permanent Swamp Forest (PSF), Semi Deciduous Transition Forest (SDTF), Shrub Savannah (SS) and Urban (U). Figure 1b shows the location of the training / control samples sites identified for each class, overlaid on a color composite image of GLC2000. The size of each site varies slightly, with about 20 pixels on average assigned for each class. The data was divided in 5 subsets to be used separately as training and control samples.



(a) (b)

Figure 1: (a) Location of the study area in Brazil. (b) GLC2000 color composite with the location of training sites for the 13 land cover classes tested in Rondonia.

Most sites were identified in the central region of Rondonia, where in the last three decades there has been considerable change in land use / land cover, mostly due to deforestation for crop and grassland production (Eva *et al.,* 2004). There are also some areas of transition between forest and regeneration, which were deforested and later abandoned, thus starting a natural regeneration process. The classes were chosen to include native vegetation, regeneration areas, intensive agriculture, flooded vegetation and urban areas.

The state of Rondonia, as well as all Brazilian Amazonia, only has two climate seasons, which are well defined – dry season and rainy season. Regarding air temperature, there is no significant variation throughout the year. Thus the main climatic influence for vegetation development is the annual hydric regime. The start of the dry season may vary from year to year, but it always includes the months of July and August.

2.2. Satellite imagery and data processing

The VGT sensor onboard SPOT4 satellite provides daily coverage of the entire Earth since 1998, at a spatial resolution of 1 km (VITO, 2012). The Normalized Difference Vegetation Index (NDVI) was computed from the extracted pixel values as :

$$\mathrm{NDVI} = (\rho_{\mathrm{NIR}} - \rho_{\mathrm{red}}) / (\rho_{\mathrm{NIR}} + \rho_{\mathrm{red}}) \quad (1)$$

where ρ_{NIR} is the reflectance at the near infrared wavelength band and ρ_{red} is the reflectance at the red waveband. The width of the red and near infrared NIR bands in the VGT sensor correspond to 0.61 – 0.68 µm and 0.78 – 0.89 µm, respectively.

Ten-day NDVI synthesis images (NDVI S10) are available from VGT globally, divided in 10 regions, being one of them South America. This synthesis images are Maximum Value Composites (MVCs) of daily NDVI images, intended to reduce the noise caused by a variety of biophysical factors. For each year, a total of 36 VGT S10 NDVI images are thus available (3 for each month). For each one of the 13 classes, the NDVI median value was computed using the training areas defined. A NDVI annual profile (time-series) was created for each class, from the 36 NDVI median values extracted from each year.

Although the NDVI S10 images are already smoothed versions of the daily NDVI images, there is still considerable noise in the temporal series data, which limits its potential for extracting meaningful information. This is particularly true for the Amazon region due to the presence of cloud coverage during long periods. A Savitzky-Golay (SG) filter (Press *et al.,* 2007) was applied to remove rare events, such as rapid transitions or other anomalies in the temporal series. The SG filter coefficients are derived by performing an unweighted linear least square fit using a polynomial of 1st degree and frame size 5.

2.3. Image classification procedure

The land cover classification of a pixel makes use of its NDVI annual profile (time-series). The profile is compared with the profiles available for each class from training, using a distance measure. In this work, two different similarity measures were used, following the experiment carried out in (Rodrigues *et al., submitted*). The first one is a rigid measure – Jaccard coefficient (Egghe, 2009), and the second one is a flexible measure – Dynamic Time Warping or DTW (Chu *et al.,* 2002), capable of incorporating shifts or distortions in time.

One useful feature to interpret land cover classified data is to have a hierarchical structure of the classes, so that simplified versions with fewer classes can be created. A hierarchical structure of the 13 initial classes was established, using the same metrics (Jaccard and DTW).

In order to evaluate the classification results, a cross-validation approach was used. The available reference data for each class was split randomly in 5 groups. These groups were used as training (80%) and control (20%) samples. The process was repeated 5 times, leaving a

different subset for validation at each time. The classification accuracy was determined for each class, as the average of the 5 classification experiments.

Using the year 2000 as reference, the NDVI profiles for the years 1999 and 2010 were processed to evaluate the influence of temporal variability in the land cover results. For 1999 are expected fewer changes in the land cover, and the eventual differences between the reference year and the studied year can be due mostly to variability in the weather and crop conditions. In a more distant year, 2010, the land cover changes are expected to be higher and, consequently, the classification process will have to deal with weather and crop variability, as well as land cover changes.

The same training areas (same pixel/geographical locations) used for 2000 were selected for 1999 and 2010. Two different approaches were considered: in the first, the same procedure described for the year 2000 was conducted in 1999 and 2010. This approach assumes that the land cover in the training areas did not suffer changes along the years. This assumption is obviously weak, particularly for the year 2010. To solve these flaws a second approach was considered, which uses training samples from the year 2000 and validation samples from the test year (1999 or 2010). This second approach allows the elimination of problems caused by the differences in annual land cover in training areas, but is vulnerable to changes associated with atmospheric conditions of the various years.

3. RESULTS AND DISCUSSION

Ten-day NDVI synthesis images from VGT were processed for Rondonia, for the years 1999, 2000 and 2010. As precipitation has considerable influence on the vegetation development, the most relevant period for this region is from April to September. Furthermore, the presence of clouds limits the use of passive satellite data from the rainy season.

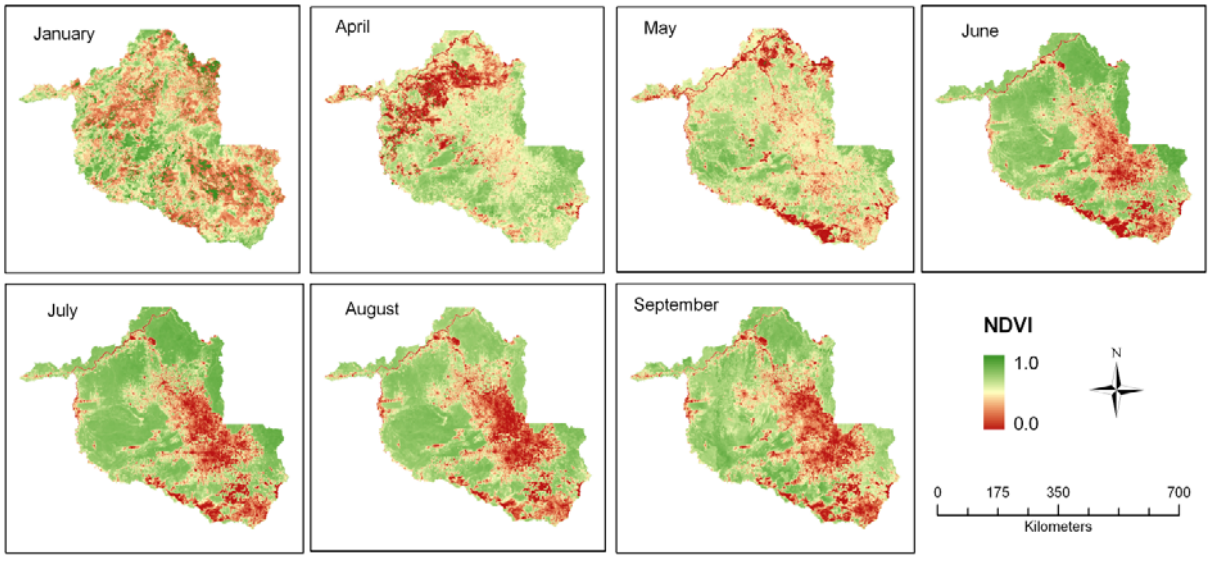
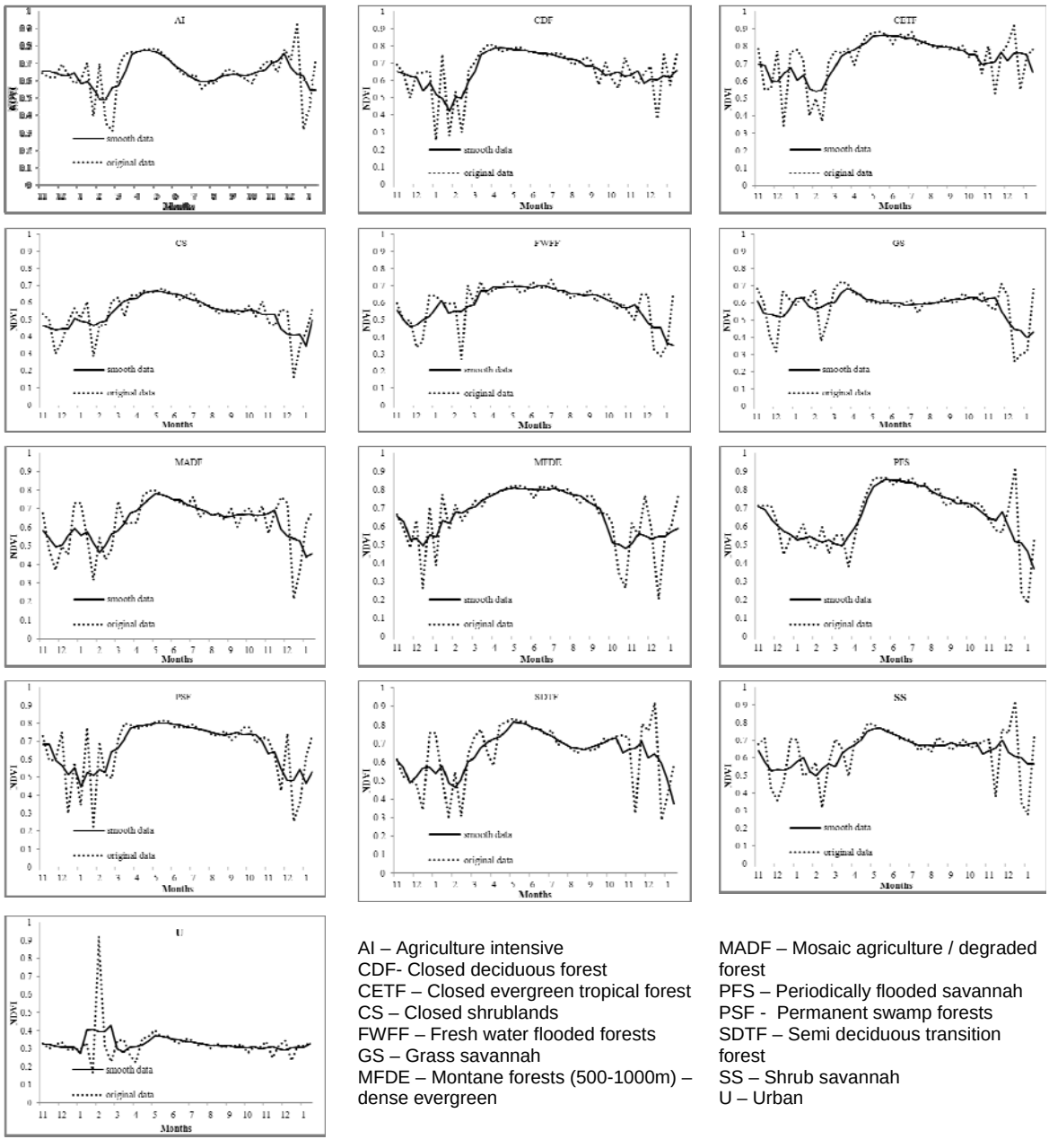


Figure 2. NDVI 10-day composites for Rondonia produced from VEGETATION data from 2000.

Mapping of the first decade NDVI value for each LCC for January and April-September period for the year 2000 are presented in figure 2. For January, due to the frequent cloud cover, the NDVI values might not be clearly related to the vegetation type (land cover class) and condition.

In the remaining images presented in figure 2, it is noticeable that the central areas of Rondonia have lower NDVI values. This is due to the fact that this part of the state has low height herbaceous or semi herbaceous vegetation type, as the soil is mostly used for agriculture and pasture.

The temporal profiles for the 13 land cover classes are presented in figure 3 for 2000, both in their original form and in a smoothed form, after the application of a SG filter.



AI – Agriculture intensive
CDF- Closed deciduous forest
CETF – Closed evergreen tropical forest
CS – Closed shrublands
FWFF – Fresh water flooded forests
GS – Grass savannah
MFDE – Montane forests (500-1000m) – dense evergreen
MADF – Mosaic agriculture / degraded forest
PFS – Periodically flooded savannah
PSF - Permanent swamp forests
SDTF – Semi deciduous transition forest
SS – Shrub savannah
U – Urban

Figure 3. NDVI time series (original and smoothed data).of the 13 land cover classes for 2000

Each LCC multi-temporal NDVI signature was consistent with its phenological characteristics and most LCC were spectrally separable at some point during the growing season. The non-

smoothed NDVI profiles show consistently a major influence of the presence of clouds especially from November to February.

Classes with smaller vegetation are the most sensitive to the hydrological regime. The class AI, shows a decrease NDVI value from June until August, a period that is considered quite dry (figure 3). The vegetation with larger size (e.g. CETF), maintain NDVI values above 0.8 during the period May-September. These higher NDVI values during the dry season could be related with the ability of vegetation to capture water in greater depth.

The temporal profiles associated with the various land cover types used show that there is a considerable similarity between a number of classes (figure 3.). It is therefore reasonable to merge some classes, thus reducing the total number of classes in the land cover map. In order to carry out this task in a suitable manner, a hierarchical structuring of the classes was established, using the Jaccard and DTW metrics.

Figure 4 shows the dendogram for the hierarchical structure of the 13 land cover classes, based on the VGTdata from 2000, using the Jaccard coefficient. Through the dendogram analysis, it is possible to identify 5 groups of classes. The first group (G1), comprising the classes 3, 4, 2, 6 and 8, represents a vegetation type that verifies a decline in the NDVI value around the final spring season, which is related to the characteristics of degraded or deciduous vegetation. The second group (G2) that includes the class 5, is a specific vegetation type that can verify a double growth-season in the same year. The third group (G3), comprising the classes 12, 13 and 10, presents a regular profile during all summer season with an accentuated NDVI peak (around 0.8) on May. The fourth group (G4), with classes 7, 11 and 1, is characterized by a high NDVI profile, representing permanent and dense forests. Lastly, the fifth group (G5), composed by class 9, presents a low and almost linear NDVI throughout all year, this being a characteristic of non-vegetated land cover classes.

The VGT data was classified using two different approaches, both with some limitations. The first one assumes that there is no difference between the land cover types present in the areas identified as training, for the period 1999-2010. It used training temporal profiles created from each year being classified. The second one uses the temporal profiles identified as training in 2000, for the classification of new observations (1999 and 2010). This approach does not account for the variability in vegetation condition due to the changing hydrological pattern.

Table 1 presents the average classification accuracy obtained for all years, at each level of aggregation, using training and control samples of the same year. In the first level of aggregation (13 classes), the accuracy obtained is greater than 80% for all years. To obtain a value of 85%, on 1999 and 2000, the number of classes must be reduced to 10, and for 2010 this number must be reduced at least 2 levels.

The results after the eighth level of aggregation (6 classes) stabilize and the aggregation process becomes time independent. According to the dendogram classes 9, 1, 11 and 7 are not yet aggregated; however the other classes were aggregated in two groups (figure 4): one includes the G1 and G2, and the other the G3. These two groups are more susceptible to suffer changes over time due to weather, soil and human conditions, which might cause an early or later leaf fall on forest vegetation types.

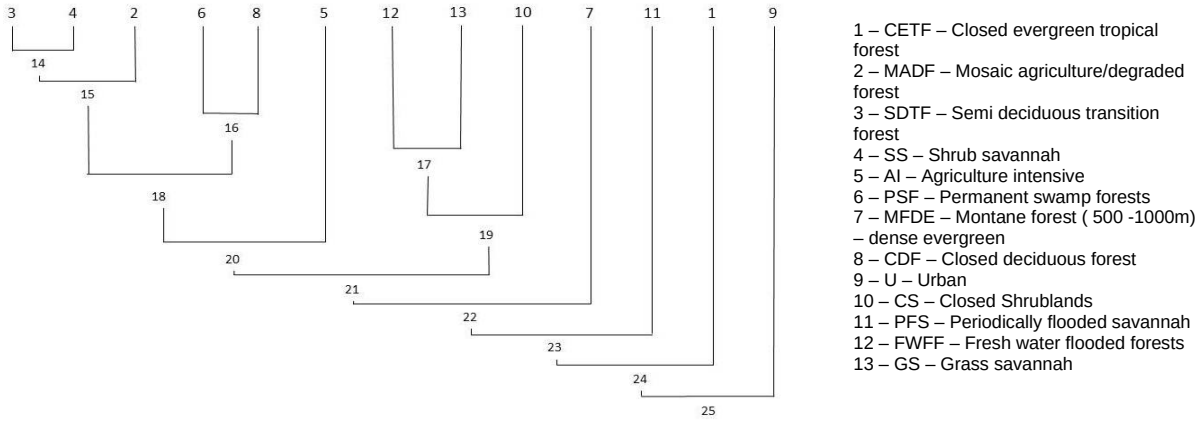


Figure 4: Dendogram for the hierarchical aggregation of the 13 land cover classes, based on the VGT data from 2000, using the Jaccard coefficient.

Table 1 – Average classification accuracy (case 1) for the 12 hierarchical levels, using the Jaccard as classifier

Year	Aggregation level											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
2000	80.5%	81.3%	82.7%	84.8%	87.6%	87.4%	92.5%	91.5%	95.1%	97.7%	96.7%	99.5%
1999	83.2%	84.2%	86.5%	84.6%	87.5%	86.7%	86.3%	88.9%	87.3%	89.1%	92.2%	99.5%
2010	80.5%	80.6%	81.2%	80.7%	82.9%	85.8%	92.0%	94.4%	94.7%	94.4%	93.5%	99.5%

Table 2 presents the average classification accuracy obtained for 1999 and 2010, at each level of aggregation, using training samples of 2000. For the first level (13 classes) the average accuracy was not higher than 41%, which shows that an error around 60% was obtained in the classification process. On both years, the average accuracy increases over the hierarchical levels. However, for 2010 the accuracy values are lower than those obtained for 1999, proving that a more distant year than the reference year (2000) is more susceptible to suffer changes.

Table 2 – Average classification accuracy (case 2) for the 12 hierarchical levels, using the Jaccard as classifier

Year	Aggregation Level											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1999	41.2%	46.9%	49.1%	53.8%	55.5%	73.0%	82.1%	86.1%	92.6%	90.4%	90.5%	99.2%
2010	31.5%	34.6%	39.3%	44.0%	50.0%	51.7%	53.8%	57.9%	62.6%	64.6%	82.2%	99.1%

Comparing with table 1, where an accuracy of 80% is achieved at the first level, in this case to obtain this precision for the year 1999, the classes need to be aggregated until the 7th hierarchical level. Thus the classes 5, 7, 11, 1 and 9 are not yet aggregated and the others classes were aggregated in the previously identified as G1 and G3 groups. For the year 2010, a precision of 80% was obtained only in the 11st level, maintaining the classes with more distinct profiles (1 and 9) as the individual classes.

CONCLUSIONS

This study presents an approach for continuous update land cover mapping using moderate spatial resolution time series data sets. The comparison of classification results over several years not only indicates the method's consistency, but also its potential to detect land cover change in tropical regions were clouds cover is frequent in the rainy season. The accuracies were obtained in relation to GLC 2000 dataset which have its own inaccuracy, thus independent ground truth data are needed to declare this approach robust and accurate.

ACKNOWLEDGMENTS

We acknowledge SPOT-image and the Flemish Institute for Technological Research (VITO, Belgium) for granting and access to SPOT-VEGETATION data. This work was developed at *Centro de Investigação em Ciências Geoespacias, Universidade do Porto* under the training of Deise Furlan funded by a Santander Mobility and Fapesp (2010/02228-0) Grants between *Universidade do Porto* and *Universidade de São Paulo*.

REFERENCES

ANA - *Agência Nacional De Águas* (Brasil). sistemas de informações hidrológicas:http://hidroweb.ana.gov.br/HidroWeb.(Accessed 7Mars 2012).

Carreiras, J., Pereira, J., Shimabukuro, Y., Stroppiana, D. 2003. Evaluation of compositing algorithms over the Brazilian Amazon using SPOT-4 Vegetation data. *International Journal of Remote Sensing*, Vol. 24, No. 17, pp. 3427-3440.

Chu, S., Keogh, E., Hart, D., Pazzani, M.,. Iterative Deepening Dynamic Time Warping for Time Series. In *Proc 2nd SIAM International Conference on Data Mining*. 2002, Maebashi City, Japan.

Egghe, L., 2009. New relations between similarity measures for vectors based on vector norms. *J. Am. Soc. Inform. Sc. Tech*. Vol. 60, No. 2, pp. 232-239.

Eva, H. D., Belward, A. S., Miranda, E. E. de, Bella, C. M. di, Gond, V., Huber, O., Jones, S., Sgrenzaroli, M., Fritz, S. 2004. "A land cover map of South America," *Global Change Biol*., Vol. 10, pp. 731–744.

Hughton, A. Skole, L., Nobre, A., Hackler, L., Lawrence, T. and Chomentowski, H. 2000. Annual fluxes of carbon from deforestation and regrowth in Brazilian Amazon. *Nature*, Vol. 403, pp. 301-304.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data*. Int. J. Remote Sens*. Vol. 21, pp. 1303-1330.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. *Numerical Recipes: The Art of Scientific Computing*, pp. 766-772.

Rodrigues, A., Marçal, A.R.S., Cunha, M. 2011. Land Cover Map update using NDVI time-series from SPOT VEGETATION – The African Continent. *International Journal of Applied Earth Observation and Geoinformation*. (*submitted*)

VITO *(Free SPOT VEGETATION products)*, 2012. http://free.vgt.vito.be/ (Accessed 20Feb2012).

Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sens. Environ.* Vol.112, pp. 1096-1116.